

GenCore version 5.1.8  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: May 9, 2006, 10:53:35 ; Search time 14 Seconds  
(without alignments)  
2219.858 Million cell updates/sec

Title: US-10-634-905-2  
Perfect score: 1705  
Sequence: 1 MKSALTRFFILLPWILVI.....EKVNLANEKPKYHLDTVKIEV 323

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues  
Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR 80.\*  
1: pir1.\*  
2: pir2.\*  
3: pir3.\*  
4: pir4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1526.5	89.5	324	2 JG0163	glucuronyltransfer
2	777.5	45.6	347	2 JC7828	glucuronyltransfer
3	529.5	31.1	356	2 T27733	hypothetical prote
4	365	21.4	259	2 T20205	hypothetical prote
5	361	21.2	290	2 T24926	hypothetical prote
6	355	20.8	248	2 T20027	hypothetical prote
7	347.5	20.4	226	2 T20447	hypothetical prote
8	336	19.7	325	2 T24737	hypothetical prote
9	180	10.6	351	2 D84788	hypothetical prote
10	169	9.9	544	2 F85435	UDP-glucuronyltran
11	111	6.5	666	2 B70803	hypothetical prote
12	107	6.3	516	2 T49422	RAD57 related prot
13	102	6.0	786	1 A47547	serine proteinase
14	101	5.9	92	2 T29701	hypothetical prote
15	99.5	5.8	813	1 S33261	protein kinase lin
16	97	5.7	586	2 T29695	hypothetical prote
17	96	5.6	767	2 S41479	DNA-binding protei
18	96	5.6	2282	2 T42717	DNA-binding protei
19	94	5.5	1360	2 T32833	hypothetical prote
20	92	5.4	497	2 S22708	homeotic protein e
21	92	5.4	1396	2 A44453	translation initia
22	91.5	5.4	367	2 H83088	membrane-bound lyt
23	91.5	5.4	445	2 A75376	probable oligosacc
24	91	5.3	1466	2 T32422	hypothetical prote
25	90.5	5.3	285	2 S08491	hypothetical prote
26	90.5	5.3	520	1 F0LJGL	gag polyprotein -
27	90.5	5.3	1520	2 T00273	hypothetical prote
28	90.5	5.3	4135	2 T42629	tenascin-X - bovin
29	89.5	5.2	366	2 AG2060	hypothetical prote

30	89.5	5.2	497	2 F83634	hypothetical prote
31	89.5	5.2	591	2 G96734	spore coat protein
32	89	5.2	428	2 A83005	conserved hypotet
33	88.5	5.2	448	2 D87146	conserved hypotet
34	88.5	5.2	818	2 A59433	KIAA0672 protein l
35	88	5.2	727	2 AD1868	hypothetical prote
36	87.5	5.1	364	2 C87455	alanine racemase l
37	87	5.1	1258	2 JC5765	inositol polyphos
38	86	5.0	964	2 D59404	plectin isoform p1
39	85.5	5.0	370	2 AG0359	probable membrane-
40	85.5	5.0	1091	2 S33596	protein-tyrosine k
41	85.5	5.0	1892	2 T18314	hypothetical prote
42	85.5	5.0	4687	1 A39638	plectin - rat
43	85	5.0	1377	2 C65159	rhaA protein precu
44	84.5	5.0	147	2 S37485	gene mgl1 protein
45	84.5	5.0	264	2 AH2041	hypothetical prote

ALIGNMENTS

RESULT 1

JG0163  
glucuronyltransferase - rat  
C;Species: Rattus norvegicus (Norway rat)  
C;Date: 23-Jul-1999 #sequence\_revision 23-Jul-1999 #text\_change 09-Jul-2004  
C;Accession: JG0163

R;Seiki, T.; Oka, S.; Terayama, K.; Imiya, K.; Kawasaki, T.  
Biochem. Biophys. Res. Commun. 255, 182-187, 1999  
A;Title: Molecular cloning and expression of a second glucuronyltransferase involved in t  
A;Reference number: JG0163; MUID:99185317; PMID:10082676  
A;Accession: JG0163  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-324 <SEI>  
A;Cross-references: UNIPROT:Q9Z137; UNIPARC:UPI000012670D; DDBJ:AB010441; NID:94519213;

Query Match 89.5%; Score 1526.5; DB 2; Length 324;  
Best Local Similarity 89.8%; Pred. No. 5.8e-119;  
Matches 291; Conservative 10; Mismatches 22; Indels 1; Gaps 1;

Qy	1	MKSALTRFFILLPWILVIIMLDVDRTRPVPLTPRPYFSPYAVGGRGARLPLRRGGPA	60
Db	1	MKSALCNRRFFILLPWILVIIMLDVDRPRRPAQLTSRPYFSPHTVCGGSRVPLRRSSPG	60
Qy	61	H-GTQXGNSRQPOPEPOLPTIYAITPTYSRPVQKAELTRLANTRQVAQLHWILVEDA	119
Db	61	RDAAEKRNESRPLQPEPLPTIYAITPTYSRPVQKAELTRLANTRQVAQLHWILVEDR	120
Qy	120	AARSELVSRFLARAGLPSTHLVPTPRRYKRPGLPRATEORNAGLAWLRQHQRQAOPG	179
Db	121	ATRSELVSSFLARAGLPNTHLVPTPRRYKRPGLPRATEORNAGLAWLRQHQRQAOPG	180
Qy	180	VLFFADDDNTYSLELFPQEMRTTRKVSVPVGLVGGRRYRPLRVENGKVGWYTGWRADRP	239
Db	181	VLFFADDDNTYSLELFPQEMRTTRKVSVPVGLVGGRRYRPLRVENGKVGWYTGWRDRP	240
Qy	240	FAIDMAGFAVSLQVILSNPKAVFKRRSQGQESDFLQITTVEELEPKANNCTKVLVW	299
Db	241	FAIDMAGFAVSLQVILSNPKAVFKRRSQGQESDFLQITTVDSLEPKANNCTKVLVW	300
Qy	300	HTRTEKVNLANEPKYHLDTVKIEV	323
Db	301	HTRTEKVNLANEPKYHMDTVNIEV	324

RESULT 2

JC7828  
glucuronyltransferase-P, long form - mouse  
C;Species: Mus musculus (house mouse)  
C;Date: 03-Jun-2002 #sequence\_revision 03-Jun-2002 #text\_change 09-Jul-2004  
C;Accession: JC7828  
R;Yamamoto, S.; Oka, S.; Saito-Ohara, F.; Inazawa, J.; Kawasaki, T.

J. Biochem. 131, 337-347, 2002  
A;Title: Molecular cloning and genomic analysis of mouse glucuronyltransferase involved  
A;Reference number: JC7828; PMID:11872162; MUID:21861939  
A;Contents: Brain  
A;Accession: JC7828  
A;Molecule type: mRNA  
A;Residues: 1-347 <YAM>  
A;Cross-references: UNIPROT:Q8R531; UNIPARC:UPI00000888CC; DDBJ:AB055781  
C;Comment: This enzyme, highly conserved during phylogenesis, is a key enzyme involved in  
C;Genetics:  
A;Gene: GICAT-P  
A;Map position: A4-9  
A;Introns: 2/1; 51/1; 220/3; 319/3

Query Match 45.6%; Score 777.5; DB 2; Length 347;  
Best Local Similarity 48.3%; Pred. No. 7.5e-57;  
Matches 167; Conservative 36; Mismatches 86; Indels 57; Gaps 8;

QY 11 ILLPWLIIII-----MLDV-----DTRRPVPP-ITPRPYF-----SPYA 44  
DB 26 IVLPTLLITVHQSLLAPLAVHKDEGSDPRHEAPPADPREYCMDSRDRIVEVVRTEYV 85  
QY 45 VGRGGARLPLRRGGPAHGQTKRNSRPPQPEPOLPTIATPTYSRPVQKAELTRIANT 104  
DB 86 -----YTRPPWSD-TLFTIHVVPTYSRPVQKAELTRIANT 121  
QY 105 FRQAQLHWILVEDAAARSELVSRFLARAGLPSTHLVPTPRYK-----RPLGPLRATE 158  
DB 122 LLHVPNLHWLVEDAPRRTPPLTARLLRDTGLNTHLVETPRNYKLRGDARDPRIPRGTM 181  
QY 159 ORNAGLAWLRQHQHQAQGVLPFADDNDTYSLELFQEMRTTRKVSVMVPGVLGGRRYE 218  
DB 182 ORNLALRWLRETFFPRNSTQPGVVYFADDNDTYSLELFEEEMRSTRVSVMVPAFVGLRYE 241  
QY 219 RPLVEN-GKVVGYTGWRAADRPFAIDMAGFAVSLQVILSNPKAVFKRGSQPGQMOSDFL 277  
DB 242 AFRVNGAGKVVGHKTVDHRPFAIDMAGFAVNLRLILQRSQAYFKLRGVKGQYQESSLL 301  
QY 278 KQITVVEELPKANNCTKVLVWHTRTEKVNLANEPKYHLDTVKIEV 323  
DB 302 RELVTLNDLEPKAANCTKILVWHTRTEKVLVNEGKGGTDPSEVI 347

RESULT 3  
T27733  
hypothetical protein ZK1307.5 - Caenorhabditis elegans  
C;Species: Caenorhabditis elegans  
C;Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 09-Jul-2004  
C;Accession: T27733  
R;Dobson, M.  
submitted to the EMBL Data Library, January 1995  
A;Reference number: Z20412  
A;Accession: T27733  
A;Status: preliminary; translated from GB/EMBL/DDBJ  
A;Molecule type: DNA  
A;Residues: 1-356 <WIL>  
A;Cross-references: UNIPROT:Q09363; UNIPARC:UPI0000135EA9; EMBL:Z47358; PIDN:CAA87436.1;  
A;Experimental source: clone ZK1307  
C;Genetics:  
A;Gene: CESP-ZK1307.5  
A;Map position: 2  
A;Introns: 46/3; 81/2; 247/1

Query Match 31.1%; Score 529.5; DB 2; Length 356;  
Best Local Similarity 44.4%; Pred. No. 2.8e-36;  
Matches 114; Conservative 41; Mismatches 91; Indels 11; Gaps 6;

QY 63 TQRNQSRRPQPEPOLPTIATPTYSRPVQKAELTRIANTFRQAQLHWILVEDAAAR 122  
DB 80 TQTRDLSLLPRVNRSTPTFTYFTTHFAAQRADLTRLSYTLSHVFNLIHWIWEDESL 139  
QY 123 SELVSRFLARAGLPSTHLVPTP-----RYKRG--LPRATEORNAGLAWLRQHQHQA 176

J. Biochem. 131, 337-347, 2002  
A;Title: Molecular cloning and genomic analysis of mouse glucuronyltransferase involved  
A;Reference number: JC7828; PMID:11872162; MUID:21861939  
A;Contents: Brain  
A;Accession: JC7828  
A;Molecule type: mRNA  
A;Residues: 1-347 <YAM>  
A;Cross-references: UNIPROT:Q8R531; UNIPARC:UPI00000888CC; DDBJ:AB055781  
C;Comment: This enzyme, highly conserved during phylogenesis, is a key enzyme involved in  
C;Genetics:  
A;Gene: GICAT-P  
A;Map position: A4-9  
A;Introns: 2/1; 51/1; 220/3; 319/3

Query Match 45.6%; Score 777.5; DB 2; Length 347;  
Best Local Similarity 48.3%; Pred. No. 7.5e-57;  
Matches 167; Conservative 36; Mismatches 86; Indels 57; Gaps 8;

QY 11 ILLPWLIIII-----MLDV-----DTRRPVPP-ITPRPYF-----SPYA 44  
DB 26 IVLPTLLITVHQSLLAPLAVHKDEGSDPRHEAPPADPREYCMDSRDRIVEVVRTEYV 85  
QY 45 VGRGGARLPLRRGGPAHGQTKRNSRPPQPEPOLPTIATPTYSRPVQKAELTRIANT 104  
DB 86 -----YTRPPWSD-TLFTIHVVPTYSRPVQKAELTRIANT 121  
QY 105 FRQAQLHWILVEDAAARSELVSRFLARAGLPSTHLVPTPRYK-----RPLGPLRATE 158  
DB 122 LLHVPNLHWLVEDAPRRTPPLTARLLRDTGLNTHLVETPRNYKLRGDARDPRIPRGTM 181  
QY 159 ORNAGLAWLRQHQHQAQGVLPFADDNDTYSLELFQEMRTTRKVSVMVPGVLGGRRYE 218  
DB 182 ORNLALRWLRETFFPRNSTQPGVVYFADDNDTYSLELFEEEMRSTRVSVMVPAFVGLRYE 241  
QY 219 RPLVEN-GKVVGYTGWRAADRPFAIDMAGFAVSLQVILSNPKAVFKRGSQPGQMOSDFL 277  
DB 242 AFRVNGAGKVVGHKTVDHRPFAIDMAGFAVNLRLILQRSQAYFKLRGVKGQYQESSLL 301  
QY 278 KQITVVEELPKANNCTKVLVWHTRTEKVNLANEPKYHLDTVKIEV 323  
DB 302 RELVTLNDLEPKAANCTKILVWHTRTEKVLVNEGKGGTDPSEVI 347

RESULT 3  
T27733  
hypothetical protein ZK1307.5 - Caenorhabditis elegans  
C;Species: Caenorhabditis elegans  
C;Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 09-Jul-2004  
C;Accession: T27733  
R;Dobson, M.  
submitted to the EMBL Data Library, January 1995  
A;Reference number: Z20412  
A;Accession: T27733  
A;Status: preliminary; translated from GB/EMBL/DDBJ  
A;Molecule type: DNA  
A;Residues: 1-356 <WIL>  
A;Cross-references: UNIPROT:Q09363; UNIPARC:UPI0000135EA9; EMBL:Z47358; PIDN:CAA87436.1;  
A;Experimental source: clone ZK1307  
C;Genetics:  
A;Gene: CESP-ZK1307.5  
A;Map position: 2  
A;Introns: 46/3; 81/2; 247/1

Query Match 31.1%; Score 529.5; DB 2; Length 356;  
Best Local Similarity 44.4%; Pred. No. 2.8e-36;  
Matches 114; Conservative 41; Mismatches 91; Indels 11; Gaps 6;

QY 63 TQRNQSRRPQPEPOLPTIATPTYSRPVQKAELTRIANTFRQAQLHWILVEDAAAR 122  
DB 80 TQTRDLSLLPRVNRSTPTFTYFTTHFAAQRADLTRLSYTLSHVFNLIHWIWEDESL 139  
QY 123 SELVSRFLARAGLPSTHLVPTP-----RYKRG--LPRATEORNAGLAWLRQHQHQA 176

DB 140 TPSIAGILKRSKIPNTHLNARTPSQKMYDDDPNWTLPRGVEQRNALLWI--QNOLSGV 197  
QY 177 QGVLPFFADDDNTYSLELFQEMRTTRKVSVMVPGVLGGRRYERPLVE-NGKVVGVWYTGHR 235  
DB 198 KEGVVYFGDDNTYDLKIFGEMRKVKNAGVWPVGVMFVETPILEKNKSGISHFNAMVK 257  
QY 236 ADPRPAIDMAGFAVSLQVILSNPKAVFKRGSQPGQMOSDFLQKITTVB-ELEPKANNCT 294  
DB 258 PERPPFIDMAAFANISLIVLSNANALFSFDVPR-GYQESTFLENLGIHRYNMEPLAEMCT 316  
QY 295 KVLVWHTRTEKVNLANE 311  
DB 317 KYVWHTRTEKPKLSKE 333

RESULT 4  
T20205  
hypothetical protein C54C8.5 - Caenorhabditis elegans  
C;Species: Caenorhabditis elegans  
C;Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 09-Jul-2004  
C;Accession: T20205  
R;Dobson, R.  
submitted to the EMBL Data Library, November 1996  
A;Reference number: Z19237  
A;Accession: T20205  
A;Status: preliminary; translated from GB/EMBL/DDBJ  
A;Molecule type: DNA  
A;Residues: 1-259 <WIL>  
A;Cross-references: UNIPROT:O17708; UNIPARC:UPI0000079C21; EMBL:Z83102; PIDN:CAB05467.1;  
A;Experimental source: clone C54C8  
C;Genetics:  
A;Gene: CESP:C54C8.5  
A;Map position: 1  
A;Introns: 33/2; 141/1; 226/3  
C;Superfamily: Caenorhabditis elegans hypothetical protein C47F8.4

Query Match 21.4%; Score 365; DB 2; Length 259;  
Best Local Similarity 37.9%; Pred. No. 8e-23;  
Matches 92; Conservative 38; Mismatches 95; Indels 18; Gaps 9;

QY 81 TIYAJTPTYSRPVQKAELTRIANTFRQAQLHWILVEDAAARSELVSRFLARAGLPSTHL 140  
DB 14 TIIVTPTYKRIITPLPDMIRLANTLAHKNLHWIIVEDGYGIVPAVRELLEKTNLSYTM 73  
QY 141 HVPTPRYKRGPLPRATEORNAGLAWLRQ-HQHQRAQGVLPFADDDNTYSLELFQE-M 198  
DB 74 AHKTAAGYP---TRGWYQRTMALRYIRSSWSKILLREHDAVVYFADDDNAYDVRIFTDYI 129  
QY 199 RTTRKVSVMVPGVLGGRRYERPLVENGVGVWYTGWRADRPFAIDMAGFAVSLQVILSNP 258  
DB 130 RNVTILGVWAVGLVGAIVEAPKVVNHKVTGNVRWSNRRRFAVDMAGFAINLKYL-NS 188  
QY 259 KAVFK---RRGSQPGQMOSDFLQKITTVVEELPKANNCTKV---LVWHTRTE--KVNLA 309  
DB 189 DAVFGKCKERGE--GAPETCLIEDIGLDEIKFPGYDATKVQDIMVWHTKSPPEIQE 246  
QY 310 NEP 312  
DB 247 DEP 249

RESULT 5  
T24926  
hypothetical protein T15D6.7 - Caenorhabditis elegans  
C;Species: Caenorhabditis elegans  
C;Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 09-Jul-2004  
C;Accession: T24926  
R;Dobson, R.  
submitted to the EMBL Data Library, November 1996  
A;Reference number: Z19956  
A;Accession: T24926  
A;Status: preliminary; translated from GB/EMBL/DDBJ  
A;Molecule type: DNA

Dd 177 SDVFGTDCRGE--GAPETCLLEDMLKMEIDIEPFGYDATKVRDIMVWHHTKTSPPEIQ 233

Qy 309 ANEP 312  
:  
Dd 235 TDQP 238

RESULT 7  
T20447  
hypothetical protein E03H4.12 - Caenorhabditis elegans  
C;Species: Caenorhabditis elegans  
C;Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 09-Jul-2004  
C;Accession: T20447

R;Dobson, R.  
submitted to the EMBL Data Library, November 1996  
A;Reference number: Z19276  
A;Accession: T20447  
A;Status: preliminary; translated from GB/EMBL/DDBJ  
A;Molecule type: DNA  
A;Residues: 1-226 <WIL>  
A;Cross-references: UNIPROT:O17751; UNIPARC:UPI000007875E; EMBL:Z81492; PIDN:C  
A;Experimental source: clone E03H4  
C;Genetics:  
A;Gene: CESP:E03H4.12  
A;Map position: 1  
A;Introns: 110/1; 194/3  
C;Superfamily: Caenorhabditis elegans hypothetical protein C47F8.4

Query Match 20.4%; Score 347.5; DB 2; Length 226;  
Best Local Similarity 37.9%; Pred. No. 1.9e-21;  
Matches 80; Conservative 37; Mismatches 81; Indels 13; Gaps 6

Qy 101 LANTFRQVAQLHWILVEDAAARSELVSRLAPAGLPSTHLHVPTPRRYKRPGLPATBOR 160  
::: : : ||| : : | : : ||| : : | : : ||| : : | : : ||| : : | : : ||| : :  
Dd 1 MSNTLKQIKNLHWIIVIEDGESLVPVAVQNVLSRSGLPYTVVTHTAKGYPAKG---WYQR 56

Qy 161 NAGLANLRQRHQH---QRAQGVLFADDDNTYSLELFE-MRTTRKYSVWPVLGVGGRR 216  
::: : : ||| : : | : : ||| : : | : : ||| : : | : : ||| : : | : : ||| : :  
Dd 57 DMALKMLTNSSQILGNHKHGAVVVFADDNDNSYDLRFDDFIIRNVKLGLGWAVGLVGGT 116

Qy 217 YERPLVENGKVGVWTGRDRPFDAIDMAGFAVSLQVILSNPKAVFKRGSGQFMQESDF 276

Dd 117 VEAPKVENKKVTSFNVKWPSRPFAVDMAGFAINLYILRS-AAVFGPKCHSGAPETCL 175

Qy 277 LKOI-TTVEELP---KANNCCKVLVWHT 303  
::: : : ||| : : | : : ||| : : | : : ||| : : | : : ||| : :  
Dd 176 LEKLGFLDDIIQPFGEKEEDKEILVWHT 206

RESULT 8  
T24737  
hypothetical protein T09B11.1 - Caenorhabditis elegans  
C;Species: Caenorhabditis elegans  
C;Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 09-Jul-2004  
C;Accession: T24737

R;McLay, K.  
submitted to the EMBL Data Library, October 1996  
A;Reference number: Z19930  
A;Accession: T24737  
A;Status: preliminary; translated from GB/EMBL/DDBJ  
A;Molecule type: DNA  
A;Residues: 1-325 <WIL>  
A;Cross-references: UNIPROT:O02304; UNIPARC:UPI000007EC17; EMBL:Z81147; PIDN:C  
A;Experimental source: clone T09B11  
C;Genetics:  
A;Gene: CESP:T09B11.1  
A;Map position: 1  
A;Introns: 28/2; 59/2; 286/2  
C;Superfamily: Caenorhabditis elegans hypothetical protein C47F8.4

Query Match 19.7%; Score 336; DB 2; Length 325;  
Best Local Similarity 33.7%; Pred. No. 2.7e-20;  
Matches 93; Conservative 36; Mismatches 87; Indels 60; Gaps 9







GenCore version 5.1.1.8  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: May 9, 2006, 10:56:24 ; Search time 40 Seconds  
(without alignments)  
776.950 Million cell updates/sec

Title: US-10-634-905-2  
Perfect score: 323  
Sequence: 1 MKSALTRFFILLPILVI.....EKVNLANEKYLHDTVKIEV 323

Scoring table: OLIGO  
Gapop 60.0 , Gapext 60.0

Searched: 283416 seqs, 96216763 residues

Word size : 1

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

- PIR 80:\*
  - 1: pirl:\*
  - 2: pirl2:\*
  - 3: pirl3:\*
  - 4: pirl4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	47	14.6	324	2 JG0163	glucuronyltransfer
2	15	4.6	347	2 JC7828	glucuronyltransfer
3	9	2.8	257	2 D84033	ribonuclease PH rp
4	8	2.5	138	2 B27863	Ac hypothetical pr
5	8	2.5	144	1 YMSWCM	macromycin precu
6	8	2.5	356	2 T27733	hypothetical prote
7	8	2.5	427	2 T03955	probable transposa
8	8	2.5	487	1 H81109	IMP dehydrogenase
9	8	2.5	487	2 F81906	IMP dehydrogenase
10	8	2.5	807	2 T02916	hypothetical prote
11	8	2.5	839	1 TQZMCA	probable transposa
12	8	2.5	3198	2 A43426	collagen alpha 2 f
13	7	2.2	64	2 AB3560	conserved cytosoli
14	7	2.2	87	2 H64521	hypothetical prote
15	7	2.2	87	2 D71984	hypothetical prote
16	7	2.2	95	2 AD3622	transcription regu
17	7	2.2	133	2 I64784	hypothetical prote
18	7	2.2	150	2 H72603	hypothetical prote
19	7	2.2	156	2 T41316	peroxisomal membra
20	7	2.2	176	2 E69519	conserved hypothet
21	7	2.2	178	2 C38487	single-stranded DN
22	7	2.2	179	1 DDECF	threonine efflux p
23	7	2.2	206	2 AH0466	amino acid ABC tra
24	7	2.2	220	2 A81159	hypothetical prote
25	7	2.2	226	2 T20447	hypothetical prote
26	7	2.2	229	1 W4WL8	E4 protein - human
27	7	2.2	237	1 S11927	licheninase (EC 3.
28	7	2.2	238	2 E81947	probable amino aci
29	7	2.2	239	2 E89830	hypothetical prote

ALIGNMENTS

RESULT 1

JG0163  
glucuronyltransferase - rat  
C:Species: Rattus norvegicus (Norway rat)  
C:Date: 23-Jul-1999 #sequence\_revision 23-Jul-1999 #text\_change 09-Jul-2004  
C:Accession: JG0163

R:Seiki, T.; Oka, S.; Terayama, K.; Imiya, K.; Kawasaki, T.  
Biochem. Biophys. Res. Commun. 255, 182-187, 1999  
A:Title: Molecular cloning and expression of a second glucuronyltransferase involved in  
A:Reference number: JG0163; PMID:99185317; PMID:10082676  
A:Accession: JG0163  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-324 <SEI>  
A:Cross-references: UNIPROT:Q9Z137; UNIPARC:UPI000012670D; DBJ:AB010441; NID:94519213; I

Query Match 14.6%; Score 47; DB 2; Length 324;  
Best Local Similarity 100.0%; Pred. No. 5.3e-40;  
Matches 47; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 176 AQPGLVFFADDNTYSLELFQEMRTTRKVSVMVPVLGGRRYERPLV 222

Db 177 AQPGLVFFADDNTYSLELFQEMRTTRKVSVMVPVLGGRRYERPLV 223

RESULT 2

JC7828  
glucuronyltransferase-P, long form - mouse  
C:Species: Mus musculus (house mouse)  
C:Date: 03-Jun-2002 #sequence\_revision 03-Jun-2002 #text\_change 09-Jul-2004  
C:Accession: JC7828  
R:Yamamoto, S.; Oka, S.; Saito-Ohara, F.; Inazawa, J.; Kawasaki, T.  
J. Biochem. 131, 337-347, 2002  
A:Title: Molecular cloning and genomic analysis of mouse glucuronyltransferase involved in  
A:Reference number: JC7828; PMID:11872162; PMID:21861939  
A:Contents: Brain  
A:Accession: JC7828  
A:Molecule type: mRNA  
A:Residues: 1-347 <YAM>  
A:Cross-references: UNIPROT:Q8R531; UNIPARC:UPI00000888CC; DBJ:AB055781  
C:Comment: This enzyme, highly conserved during phylogenesis, is a key enzyme involved in

Query Match 4.6%; Score 15; DB 2; Length 347;  
Best Local Similarity 100.0%; Pred. No. 4.2e-07;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 86 TPTYSRPVQKAELTR 100

|||||

Db 103 TPTYSRPVQKAELTR 117

RESULT 3

ribonuclease PH rph [imported] - Bacillus halodurans (strain C-125)

C:Species: Bacillus halodurans

C>Date: 01-Dec-2000 #sequence\_revision 01-Dec-2000 #text\_change 09-Jul-2004

C:Accession: D84033

R:Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; Hira Nucleic Acids Res. 28, 4317-4331, 2000

A:Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans and A:Reference number: A83650; MUID:20512582; PMID:11058132

A:Accession: D84033

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-257 <STO>

A:Cross-references: UNIPROT:Q9K8D8; UNIPARC:UPI0000134454; GB:AP001517; GB:BA0000004; NID A:Experimental source: strain C-125

C:Genetics:

C:Superfamily: tRNA nucleotidyltransferase

Query Match 2.8%; Score 9; DB 2; Length 257;

Best Local Similarity 100.0%; Pred. No. 0.47; Mismatches 0; Indels 0; Gaps 0;

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 153 LPRATEQRN 161

Db 66 LPRATEQRN 74

|||||

RESULT 4

Ac hypothetical protein ORPa - maize (fragment)

C:Species: Zea mays (maize)

C>Date: 19-Nov-1988 #sequence\_revision 19-Nov-1988 #text\_change 30-Sep-1993

C:Accession: B27863

R:Kunze, R.; Stochaj, U.; Laufs, J.; Starlinger, P.

EMBO J. 6, 1555-1563, 1987

A:Title: Transcription of transposable element Activator (Ac) of Zea mays L.

A:Reference number: A91069

A:Accession: B27863

A:Molecule type: mRNA

A:Residues: 1-138 <RUN>

A:Cross-references: UNIPARC:UPI000017835B

C:Superfamily: hypothetical transposase Ac9

Query Match 2.5%; Score 8; DB 2; Length 138;

Best Local Similarity 100.0%; Pred. No. 2.8; Mismatches 0; Indels 0; Gaps 0;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 71 PQQPEPQ 78

Db 113 PQQPEPQ 120

|||||

RESULT 5

YMSMCM

macromycin precursor - Streptomyces "macromyceticus"

N:Alternate names: auroomycin apoprotein (AUR)

N:Contains: macromycin (MCR)

C:Species: Streptomyces "macromyceticus"

C>Date: 31-May-1979 #sequence\_revision 31-Dec-1992 #text\_change 09-Jul-2004

C:Accession: A45766; A92409; A90212; A01812; JS0338

R:Sakata, N.; Kanbe, T.; Tanabe, M.; Hayashi, H.; Hori, M.; Hotta, K.; Hamada, M.

J. Antibiot. 42, 1704-1712, 1989

A:Title: Nucleotide sequence of the macromycin apoprotein gene and its expression in S A:Reference number: A45766; MUID:90061944; PMID:2479629

A:Accession: A45766

A:Molecule type: DNA

A:Residues: 1-144 <SAK>

A:Cross-references: UNIPROT:P01549; UNIPARC:UPI000012EB0C; GB:D90006; NID:g217017; PIDN:

A:Experimental source: strain M480-M1

R:Samy, T.S.A.; Hahn, K.S.; Modest, E.J.; Lampman, G.W.; Keutmann, H.T.; Umezawa, H.; Her J. Biol. Chem. 258, 183-191, 1983

A:Title: Primary structure of macromycin, an antitumor antibiotic protein.

A:Reference number: A92409; MUID:83082859; PMID:6848492

A:Accession: A92409

A:Molecule type: protein

A:Residues: 33-110,'D',112-144 <SAM>

A:Cross-references: UNIPARC:UPI00001125AA

R:Sawyer, T.H.; Gietzow, K.; Olson, M.O.J.; Busch, H.; Prestayko, A.W.; Crooke, S.T. Biochem. Biophys. Res. Commun. 86, 1133-1138, 1979

A:Title: Amino terminal amino acid sequence of macromycin, a protein antitumor antibiotic A:Reference number: A90212; MUID:79165840; PMID:155453

A:Accession: A90212

A:Molecule type: protein

A:Residues: 33-47,'E',49-64,66,'ES',69-70,'A',72-77,'P' <SAW>

A:Cross-references: UNIPARC:UPI0000173628

R:Van Rosy, P.; Beerman, T.A. Proc. Natl. Acad. Sci. U.S.A. 86, 6587-6591, 1989

A:Title: Crystal structure analysis of auroomycin apoprotein (macromycin) shows import A:Reference number: A33726; MUID:89367294; PMID:2771945

C:Contents: annotation; X-ray crystallography, 1.5 angstroms

C:Comment: The antibiotic activity (for gram-positive bacteria) and the antitumor activit ific inhibition of DNA synthesis, with consequent arrest of mitosis, in these susceptible C:Genetics:

A:Gene: mcma

C:Superfamily: macromycin

C:Keywords: antibiotic; DNA binding; pigment binding

F:1-32/Domain: signal sequence #status predicted <SIG>

F:33-144/Product: macromycin #status experimental <MAT>

F:68-78,120-125/Disulfide bonds: #status experimental

Query Match 2.5%; Score 8; DB 1; Length 144;

Best Local Similarity 100.0%; Pred. No. 3; Mismatches 0; Indels 0; Gaps 0;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 127 SRFLARAG 134

Db 6 SRFLARAG 13

|||||

RESULT 6

T27733

hypothetical protein ZK1307.5 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C>Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 09-Jul-2004

C:Accession: T27733

R:Berks, M.

submitted to the EMBL Data Library, January 1995

A:Reference number: Z20412

A:Accession: T27733

A>Status: preliminary; translated from GB/EMBL/DBDJ

A:Molecule type: DNA

A:Residues: 1-356 <WIL>

A:Cross-references: UNIPROT:Q09363; UNIPARC:UPI0000135EA9; EMBL:Z47358; PIDN:CAA87436.1; A:Experimental source: clone ZK1307

C:Genetics:

A:Gene: CESP:ZK1307.5

A:Map position: 2

A:Introns: 46/3; 81/2; 247/1

Query Match 2.5%; Score 8; DB 2; Length 356;

Best Local Similarity 100.0%; Pred. No. 6.7; Mismatches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 298 VWHTRTEK 305

Db 320 VWHTRTEK 327

|||||

RESULT 7

T03955

probable transposase - maize transposable element Ac



C;Species: Zea mays (maize)  
 C;Date: 23-Apr-1999 #sequence\_revision 23-Apr-1999 #text\_change 09-Jul-2004  
 C;Accession: T03955  
 R;Kunze, R.; Starlinger, P.  
 EMBL J. 8, 3177-3185, 1989  
 A;Title: The putative transposase of transposable element Ac from Zea mays L. interacts  
 A;Reference number: Z15160; MUID:90059907; PMID:2555157  
 A;Accession: T03955  
 A;Status: preliminary; translated from GB/EMBL/DBJ  
 A;Molecule type: DNA  
 A;Residues: 1-427 <KUN>  
 A;Cross-references: UNIPROT:Q9SAP9; UNIPARC:UPI000009D1DC; EMBL:X01380; NID:g22489; PIDN  
 C;Genetics:  
 A;Mobile element: transposable element Ac  
 C;Superfamily: hypothetical transposase Ac9

Query Match 2.5%; Score 8; DB 2; Length 427;  
 Best Local Similarity 100.0%; Pred. No. 7.9;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 71 PQQPPEPQ 78  
 Db 11 PQQPPEPQ 18  
 |||||

RESULT 8  
 H81109  
 IMP dehydrogenase (EC 1.1.1.205) NM01201 [similarity] - Neisseria meningitidis (strain M  
 N;Alternate names: IMP:NAD+ oxidoreductase; IMPDH; inosine-5'-monophosphate dehydrogenase  
 C;Species: Neisseria meningitidis  
 C;Date: 05-May-2000 #sequence\_revision 05-May-2000 #text\_change 09-Jul-2004  
 C;Accession: H81109  
 R;Tetzelin, H.; Saunders, N.J.; Heidelberg, J.; Jeffries, A.C.; Nelson, K.E.; Eisen, J.A.  
 Hickey, E.K.; Haft, D.H.; Salzberg, S.L.; White, O.; Fleischmann, R.D.; Dougherty, B.A.;  
 ri, H.; Qin, H.; Vamathevan, J.; Gill, J.; Scarlato, V.; Masiagnani, V.; Pizza, M.  
 Science 287, 1809-1815, 2000  
 A;Authors: Grandi, G.; Sun, L.; Smith, H.O.; Fraser, C.M.; Moxon, E.R.; Rappuoli, R.; Ve  
 A;Title: Complete genome sequence of Neisseria meningitidis serogroup B strain MC58.  
 A;Reference number: A81000; MUID:20175755; PMID:10710307  
 A;Accession: H81109  
 A;Molecule type: DNA  
 A;Residues: 1-487 <TET>  
 A;Cross-references: UNIPROT:Q9JZB5; UNIPARC:UPI00000C4660; GB:AE002098; NID  
 A;Experimental source: serogroup B, strain MC58  
 C;Genetics:  
 A;Gene: NM01201  
 C;Function:  
 A;Description: catalyzes the oxidation of inosine 5'-phosphate by NAD and one molecule o  
 A;Pathway: purine nucleotide biosynthesis  
 C;Superfamily: inosine-5'-monophosphate dehydrogenase; CBS homology; IMP dehydrogenase a  
 C;Keywords: duplication; GMP biosynthesis; NAD; oxidoreductase; purine nucleotide biosyn  
 F;8-74/Domain: IMP dehydrogenase amino-terminal homology <IDHN>  
 F;95-144/Domain: CBS homology <CBS1>  
 F;158-206/Domain: CBS homology <CBS2>  
 F;207-451/Domain: IMP dehydrogenase catalytic homology <IDHC>  
 F;305/Active site: Cys #status predicted

Query Match 2.5%; Score 8; DB 1; Length 487;  
 Best Local Similarity 100.0%; Pred. No. 8.8;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 222 VENGKVVG 229  
 Db 125 VENGKVVG 132  
 |||||

RESULT 9  
 F81906  
 IMP dehydrogenase (EC 1.1.1.205) NM01372 [similarity] - Neisseria meningitidis (strain 2  
 C;Species: Neisseria meningitidis  
 C;Date: 05-May-2000 #sequence\_revision 05-May-2000 #text\_change 09-Jul-2004  
 C;Accession: F81906  
 R;Parkhill, J.; Achtman, M.; James, K.D.; Bentley, S.D.; Churcher, C.; Klee, S.R.; Morel

; Holroyd, S.; Jagels, K.; Leather, S.; Moule, S.; Mungall, K.; Quail, M.A.; Rajandream,  
 Nature 404, 502-506, 2000  
 A;Title: Complete DNA sequence of a serogroup A strain of Neisseria meningitidis Z2491.  
 A;Reference number: A81775; MUID:20222556; PMID:10761919  
 A;Accession: F81906  
 A;Status: preliminary  
 A;Molecule type: DNA  
 A;Residues: 1-487 <PAR>  
 A;Cross-references: UNIPROT:Q9JUD0; UNIPARC:UPI00000C4BAA; GB:AL162755; GB:AL157959; NID:  
 A;Experimental source: serogroup A, strain Z2491  
 C;Genetics:  
 A;Gene: guaB; NM01372  
 C;Superfamily: inosine-5'-monophosphate dehydrogenase; CBS homology; IMP dehydrogenase an  
 C;Keywords: GMP biosynthesis; NAD; oxidoreductase  
 F;8-74/Domain: IMP dehydrogenase amino-terminal homology <IDN>  
 F;305/Active site: Cys #status predicted

Query Match 2.5%; Score 8; DB 2; Length 487;  
 Best Local Similarity 100.0%; Pred. No. 8.8;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 222 VENGKVVG 229  
 Db 125 VENGKVVG 132  
 |||||

RESULT 10  
 T02916  
 hypothetical protein a - maize transposable element Ac  
 C;Species: Zea mays (maize)  
 C;Date: 24-Mar-1999 #sequence\_revision 24-Mar-1999 #text\_change 09-Jul-2004  
 C;Accession: T02916  
 R;Kunze, R.  
 submitted to the EMBL Data Library, February 1988  
 A;Reference number: Z14767  
 A;Accession: T02916  
 A;Status: preliminary; translated from GB/EMBL/DBJ  
 A;Molecule type: mRNA  
 A;Residues: 1-807 <KUN>  
 A;Cross-references: UNIPROT:P08770; UNIPARC:UPI000016E023; EMBL:X05424; NID:g22112; PIDN  
 C;Genetics:  
 A;Mobile element: transposable element Ac  
 A;Introns: 41/1; 490/2; 756/1; 786/3  
 C;Superfamily: hypothetical transposase Ac9

Query Match 2.5%; Score 8; DB 2; Length 807;  
 Best Local Similarity 100.0%; Pred. No. 14;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 71 PQQPPEPQ 78  
 Db 113 PQQPPEPQ 120  
 |||||

RESULT 11  
 TQ2MCA  
 probable transposase - maize transposon Ac9  
 C;Species: Zea mays (maize)  
 C;Date: 17-May-1985 #sequence\_revision 17-May-1985 #text\_change 09-Jul-2004  
 C;Accession: A03540; S18636  
 R;Pohlman, R.F.; Fedoroff, N.V.; Messing, J.  
 Cell 37, 635-643, 1984  
 A;Title: The nucleotide sequence of the maize controlling element activator.  
 A;Reference number: A90848; MUID:84205699; PMID:6327080  
 A;Accession: A03540  
 A;Molecule type: DNA  
 A;Residues: 1-839 <POH>  
 A;Cross-references: UNIPROT:P03010; UNIPARC:UPI00001372D4; GB:K01964; NID:g168402  
 R;Feldmar, S.; Kunze, R.  
 EMBL J. 10, 4003-4010, 1991  
 A;Title: The OrfA protein, the putative transposase of maize transposable element Ac, hai  
 A;Reference number: S18636; MUID:92097517; PMID:1661668  
 A;Accession: S18636

A;Status: nucleic acid sequence not shown; not compared with conceptual translation  
A;Molecule type: DNA  
A;Residues: 103-184 <FEL>  
A;Cross-references: UNIPARC:UPI0000174897  
C;Comment: This protein is coded by the transposable maize controlling element "Activator" of the maize genome.  
C;Superfamily: hypothetical transposase Ac9  
C;Keywords: DNA binding; DNA replication  
F;136-145/Region: DNA binding #status experimental

Query Match 2.5%; Score 8; DB 1; Length 839;  
Best Local Similarity 100.0%; Pred. No. 14;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 71 PQQPPEPQ 78  
Db 58 PQQPPEPQ 65  
|||||

RESULT 12  
A43426  
collagen alpha 2 fibrillar chain precursor - sea urchin (Strongylocentrotus purpuratus)  
C;Species: Strongylocentrotus purpuratus (purple urchin)  
C;Date: 04-Mar-1993 #sequence\_revision 18-Nov-1994 #text\_change 09-Jul-2004  
C;Accession: A43426  
R;Exposito, J.Y.; D'Alessio, M.; Ramirez, F.  
J. Biol. Chem. 267, 17404-17408, 1992  
A;Title: Novel amino-terminal propeptide configuration in a fibrillar procollagen undergoes a conformational change during processing  
A;Reference number: A43426; MUID:92381062; PMID:1380962  
A;Accession: A43426  
A;Status: preliminary; not compared with conceptual translation  
A;Molecule type: nucleic acid  
A;Residues: 1-3198 <EXP>  
A;Cross-references: UNIPROT:Q26639; UNIPARC:UPI0000077B9A; GB:M92041; NID:g161448; PIDN:D71984  
A;Note: sequence extracted from NCBI backbone (NCBIP:111965)  
F;48-106/Domain: von Willebrand factor type C repeat homology <VMC>  
F;2978-3198/Domain: fibrillar collagen carboxyl-terminal homology <FCC>

Query Match 2.5%; Score 8; DB 2; Length 3198;  
Best Local Similarity 100.0%; Pred. No. 48;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 208 PVGLVGGR 215  
Db 2768 PVGLVGGR 2775  
|||||

RESULT 13  
AB3560  
conserved cytosolic protein [imported] - Brucella melitensis (strain 16M)  
C;Species: Brucella melitensis  
C;Date: 01-Feb-2002 #sequence\_revision 01-Feb-2002 #text\_change 09-Jul-2004  
C;Accession: AB3560  
R;DelVecchio, V.G.; Kapral, V.; Redkar, R.J.; Patra, G.; Mujer, C.; Los, T.; Ivanova, I.; Mazur, M.; Goldsman, E.; Salkov, E.; Elzer, P.H.; Hagius, S.; O'Callaghan, D.; Letesee Proc. Natl. Acad. Sci. U.S.A. 99, 443-448, 2002  
A;Title: The genome sequence of the facultative intracellular pathogen Brucella melitensis  
A;Reference number: AD3252; PMID:11756688  
A;Accession: AB3560  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-64 <KUR>  
A;Cross-references: UNIPROT:Q8FVF2; UNIPARC:UPI0000058482; GB:AE008918;  
A;Experimental source: strain 16M  
C;Genetics:  
A;Gene: BMEI0403  
A;Map position: II

Query Match 2.2%; Score 7; DB 2; Length 64;  
Best Local Similarity 100.0%; Pred. No. 15;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 122 RSELVSR 128

Db 33 RSELVSR 39  
|||||

RESULT 14  
H64521  
hypothetical protein HP0016 - Helicobacter pylori (strain 26695)  
C;Species: Helicobacter pylori  
C;Date: 09-Aug-1997 #sequence\_revision 09-Aug-1997 #text\_change 09-Jul-2004  
C;Accession: H64521  
R;Tomb, J.F.; White, O.; Kerlavage, A.R.; Clayton, R.A.; Sutton, G.G.; Fleischmann, R.D.; Peterson, S.; Loftus, B.; Richardson, D.; Dodson, R.; Khalak, H.G.; Glodek, A.; McKenney, J.D.; Kelley, J.M.; Cotton, M.D.; Weidman, J.M.; Fujii, C.; Bowman, C.; Watthey, L. Nature 388, 539-547, 1997  
A;Authors: Wallin, E.; Hayes, W.S.; Borodovsky, M.; Karpk, P.D.; Smith, H.O.; Fraser, C. A;Title: The complete genome sequence of the gastric pathogen Helicobacter pylori.  
A;Reference number: A64520; MUID:97394467; PMID:9252185  
A;Accession: H64521  
A;Status: preliminary; nucleic acid sequence not shown; translation not shown  
A;Molecule type: DNA  
A;Residues: 1-87 <TOM>  
A;Cross-references: UNIPROT:Q24861; UNIPARC:UPI00000C07CC; GB:AE000524; GB:AE000511; NID:Q24861

Query Match 2.2%; Score 7; DB 2; Length 87;  
Best Local Similarity 100.0%; Pred. No. 20;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 FILLPWI 16  
Db 36 FILLPWI 42  
|||||

RESULT 15  
D71984  
hypothetical protein jhp0014 - Helicobacter pylori (strain J99)  
C;Species: Helicobacter pylori  
A;Variety: strain J99  
C;Date: 12-Feb-1999 #sequence\_revision 12-Feb-1999 #text\_change 09-Jul-2004  
C;Accession: D71984  
R;Alm, R.A.; Ling, L.S.L.; Moir, D.T.; King, B.L.; Brown, E.D.; Doig, P.C.; Smith, D.R.; Ives, C.; Gibson, R.; Merberg, D.; Mills, S.D.; Jiang, Q.; Taylor, D.E.; Vovis, G.F.; Nature 397, 176-180, 1999  
A;Title: Genomic sequence comparison of two unrelated isolates of the human gastric pathogen Helicobacter pylori  
A;Reference number: A71800; MUID:99120557; PMID:9923682  
A;Accession: D71984  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-87 <ARN>  
A;Cross-references: UNIPROT:Q9ZM45; UNIPARC:UPI00000D713D; GB:AE001441; GB:AE001439; NID:Q9ZM45  
A;Experimental source: strain J99  
C;Genetics:  
A;Gene: jhp0014

Query Match 2.2%; Score 7; DB 2; Length 87;  
Best Local Similarity 100.0%; Pred. No. 20;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 FILLPWI 16  
Db 36 FILLPWI 42  
|||||

Search completed: May 9, 2006, 10:59:43  
Job time : 42 secs

GenCore version 5.1.8  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: May 9, 2006, 10:56:10 ; Search time 59 Seconds  
(without alignments)

3862.469 Million cell updates/sec

Title: US-10-634-905-2

Perfect score: 323

Sequence: 1 MKSALETRFFILLFWILVI.....EKVLANEPKYHLDTVKIEV 323

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 2166443 seqs, 705528306 residues

Word size : 1

Total number of hits satisfying chosen parameters: 2166443

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : UniProt 05.80.\*

1: uniprot\_sprot.\*

2: uniprot\_trembl.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	323	100.0	323	1 B3GA2_HUMAN	Q9np25 homo sapien
2	323	100.0	323	2 Q599K0_PANTR	Q599k0 pan troglod
3	236	73.1	329	1 B3GA2_CANFA	Q5ca26 canis famil
4	105	32.5	326	2 Q599K1_BOVIN	Q599k1 bos taurus
5	62	19.2	304	2 Q5CA57_CHICK	Q5ca57 gallus gall
6	54	16.7	324	1 B3GA2_MOUSE	P52770 mus musculu
7	54	16.7	349	2 Q5DTT7_MOUSE	Q5dt7 mus musculu
8	47	14.6	324	1 B3GA2_RAT	Q9z137 rattus norv
9	29	9.0	331	2 Q599J8_XENTR	Q599j8 xenopus tro
10	27	8.4	330	2 Q599J7_XENLA	Q599j7 xenopus lae
11	23	7.1	327	2 Q5CB07_BRARE	Q5cb07 brachydanio
12	22	6.8	303	2 Q599J9_ORYLA	Q599j9 oryzias lat
13	21	6.5	334	2 Q5CB02_BRARE	Q5cb02 brachydanio
14	21	6.5	335	2 Q5CB01_ORYLA	Q5cb01 oryzias lat
15	21	6.5	335	2 Q5CA29_TETNG	Q5ca29 tetraodon n
16	21	6.5	335	2 Q5CB00_FUGRU	Q5cb00 fugu rubrip
17	21	6.5	359	2 Q4SE14_TETNG	Q4se14 tetraodon n
18	20	6.2	304	2 Q5CA27_FUGRU	Q5ca27 fugu rubrip
19	20	6.2	304	2 Q5CA28_TETNG	Q5ca28 tetraodon n
20	20	6.2	316	2 Q6DBR2_BRARE	Q6db2 brachydanio
21	20	6.2	317	2 Q7T1E2_CHICK	Q7t1e2 gallus gall
22	20	6.2	413	2 Q4SG61_TETNG	Q4sg61 tetraodon n
23	19	5.9	333	2 Q599K2_BOVIN	Q599k2 bos taurus
24	15	4.6	277	2 Q8R1V2_MOUSE	Q8r1v2 mus musculu
25	15	4.6	317	2 Q7T1E1_CHICK	Q7t1e1 gallus gall
26	15	4.6	332	1 B3GA1_PANTR	Q5cb04 p galactosy
27	15	4.6	334	1 B3GA1_HUMAN	Q5p2w7 homo sapien
28	15	4.6	334	1 B3GA1_MOUSE	Q9cwf3 mus musculu
29	15	4.6	334	1 B3GA1_RAT	Q35789 r galactosy
30	15	4.6	334	2 Q96F87_HUMAN	Q96f87 homo sapien
31	15	4.6	334	2 Q6PIG8_MOUSE	Q6pig8 mus musculu

32	15	4.6	335	1 B3GA1_CANFA	Q5cb03 c galactosy
33	15	4.6	345	2 Q8BLT8_MOUSE	Q8blt8 mus musculu
34	15	4.6	347	2 Q8R531_MOUSE	Q8r531 mus musculu
35	15	4.6	394	2 Q5DTF9_MOUSE	Q5dtf9 mus musculu
36	14	4.3	334	2 Q5CB06_ORYLA	Q5cb06 oryzias lat
37	14	4.3	342	2 Q63ZM6_XENLA	Q63zm6 xenopus lae
38	13	4.0	159	2 Q8AYG7_CHICK	Q8ayg7 gallus gall
39	13	4.0	190	2 Q8ND18_HUMAN	Q8nd18 homo sapien
40	13	4.0	328	2 Q567G3_BRARE	Q567g3 brachydanio
41	13	4.0	329	2 Q8JH19_BRARE	Q8jh19 brachydanio
42	13	4.0	339	2 Q599K3_XENTR	Q599k3 xenopus tro
43	12	3.7	303	2 Q599K4_XENTR	Q599k4 xenopus tro
44	11	3.4	148	2 Q8AYG6_CHICK	Q8ayg6 gallus gall
45	11	3.4	242	2 Q7T1E2_CHICK	Q7t1e2 gallus gall

#### ALIGNMENTS

RESULT 1  
B3GA2\_HUMAN  
ID B3GA2\_HUMAN STANDARD; PRT; 323 AA.  
AC QNPZ5; Q5JS09; Q8TF38; Q96NK4;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 13-SEP-2005 (Rel. 48, Last annotation update)  
DE Galactosylgalactosylxylosylprotein 3-beta-glucuronosyltransferase 2  
DE (EC 2.4.1.135) (Beta-1,3-glucuronyltransferase 2)  
DE (Glucuronosyltransferase-S) (GICAT-S) (UDP-glucuronosyltransferase-S)  
DE (GICAT-D)  
GN Name=B3GAT2; Synonyms=GLCATS, KIAA1963;  
OS Homo sapiens (Human)  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;  
OC Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP NUCLEOTIDE SEQUENCE.  
RX MEDLINE=22410182; PubMed=12522689; DOI=10.1007/s100380200103;  
RA Marcos I., Galan J.J., Borrego S., Antinolo G.;  
RT "Cloning, characterization, and chromosome mapping of the human GICAT-S gene."  
RL J. Hum. Genet. 47:677-680 (2002).  
RN [2]  
RP NUCLEOTIDE SEQUENCE [LARGE SCALE MRNA].  
RC TISSUE=Brain;  
RX MEDLINE=21842142; PubMed=11853319;  
RA Nagase T., Kikuno R., Ohara O.;  
RT "Prediction of the coding sequences of unidentified human genes. XXII. The complete sequences of 50 new cDNA clones which code for large proteins."  
RL DNA Res. 8:319-327 (2001).  
RN [3]  
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].  
RX MEDLINE=22935763; PubMed=14574404; DOI=10.1038/nature02055;  
RA Mungall A.J., Palmer S.A., Sims S.K., Edwards C.A., Ashurst J.L., Wilming L., Jones M.C., Horton R., Hunt S.E., Scott C.E., Gilbert J.G.R., Clamp M.E., Bethel G., Milne S., Ainscough R., Almeida J.P., Ambrose K.D., Andrews T.D., Ashwell R.I.S., Babbage A.K., Bagguley C.L., Bailey J., Banerjee R., Barker D.J., Barlow K.F., Bates K., Beare D.M., Beasley H., Beasley O., Bird C.P., Blakey S.E., Bray-Allen S., Brook J., Brown A.J., Brown J.Y., Burford D.C., Burrill W., Burton J., Carder C., Carter N.P., Chapman J.C., Clark S.Y., Clark G., Clee C.M., Clegg S., Cobley V., Collier R.E., Collins J.E., Colman L.K., Corby N.R., Coville G.J., Culley K.M., Dhali P., Davies J., Dunn M., Earthrowl M.E., Ellington A.E., Evans K.A., Faulkner L., Francis M.D., Frankish A., Frankland J., French L., Garner P., Garnett J., Ghori M.J., Gilby L.M., Gillson C.J., Glithero R.J., Graham D.V., Grant M., Gribble S., Griffiths C., Griffiths M.N.D., Hall R., Halls K.S., Hammond S., Harley J.L., Hart E.A., Heath P.D., Heathcote R., Holmes S.J., Howden P.J., Howe K.L., Howell G.R., Huckle E., Humphray S.J., Humphries M.D., Hunt A.R., Johnson C.M., Joy A.A.,

RA Kay M., Keenan S.J., Kimberley A.M., King A., Laird G.K., Langford C.,  
 RA Lawlor S., Leongamornlert D.A., Leversha M., Lloyd C.R., Lloyd D.M.,  
 RA Loveland J.E., Lovell J., Martin S., Mashreghi-Mohammadi M.,  
 RA Maslen G.L., Matthews L., McCann O.T., McLaren S.J., McLeay K.,  
 RA McMurray A., Moore M.J.F., Mullikin J.C., Niblett D., Nickerson T.,  
 RA Novik K.L., Oliver K., Overton-Larty E.K., Parker A., Patel R.,  
 RA Pearce A.V., Peck A.I., Phillimore B.J.C.T., Phillips S., Plumb R.W.,  
 RA Porter K.M., Ramsey Y., Ranby S.A., Rice C.M., Ross M.T., Searle S.M.,  
 RA Sehra H.K., Sheridan E., Skuce C.D., Smith S., Smith M., Spraggon L.,  
 RA Squares S.L., Steward C.A., Sycamore N., Tamiyn-Hall G., Tester J.,  
 RA Theaker A.J., Thomas D.W., Thorpe A., Tracey A., Tromans A., Tubby B.,  
 RA Wall M., Wallis J.M., West A.P., White S.S., Whitehead S.L.,  
 RA Whittaker H., Wild A., Willey D.J., Wilmer T.E., Wood J.M., Wray P.W.,  
 RA Wyatt J.C., Young L., Younger R.M., Bentley D.R., Coulson A.,  
 RA Durbin R., Hubbard T., Sulston J.E., Dunham I., Rogers J., Beck S.;  
 RT "The DNA sequence and analysis of human chromosome 6.";  
 RL Nature 425:805-811(2003).  
 (4)  
 RP NUCLEOTIDE SEQUENCE [LARGE SCALE MRNA] OF 157-323.  
 RC TISSUE=fetal brain;  
 RX PubMed=14702039; DOI=10.1038/ng1285;  
 RA Oka T., Suzuki Y., Nishikawa T., Otsuki T., Sugiyama T., Irie R.,  
 RA Wakamatsu A., Hayashi K., Sato H., Nagai K., Kimura K., Makita H.,  
 RA Sekine M., Obayashi M., Nishi T., Shibahara T., Tanaka T., Iehi S.,  
 RA Yamamoto J.-I., Saito K., Kawai Y., Isono Y., Nakamura Y.,  
 RA Nagahara K., Murakami K., Yasuda T., Iwayanagi T., Wagatsuma M.,  
 RA Shiratori A., Sudo H., Hosoiri T., Kaku Y., Kodaira H., Kondo H.,  
 RA Sugawara M., Takahashi M., Kanda K., Yokoi T., Furuya T., Kikkawa E.,  
 RA Omura Y., Abe K., Kamihara K., Katsuma T., Sato K., Tanikawa M.,  
 RA Yamazaki M., Ninomiya K., Iehibashi T., Yamaashita H., Murakawa K.,  
 RA Fujimori K., Tanai H., Kimata M., Watanabe M., Hiraoka S., Chiba Y.,  
 RA Ishida S., Ono Y., Takiguchi S., Watanabe S., Yosida M., Hotuta T.,  
 RA Kusano J., Kanehori K., Takahashi-Fujii A., Hara H., Tanase T.-O.,  
 RA Nomura Y., Togiya S., Komai F., Hara R., Takeuchi K., Arita M.,  
 RA Imose N., Musashino K., Yuuki H., Oshima A., Sasaki N., Aotsuka S.,  
 RA Yoshikawa Y., Matsunawa H., Ichihara T., Shiohata N., Sano S.,  
 RA Moriya S., Momiyama H., Satoh N., Takami S., Terashima Y., Suzuki O.,  
 RA Nakagawa S., Senoh A., Mizoguchi H., Goto Y., Shimizu F., Wakebe H.,  
 RA Hishigaki H., Watanabe T., Sugiyama A., Takemoto M., Kawakami B.,  
 RA Yamazaki M., Watanabe K., Kumagai A., Itakura S., Fukuzumi Y.,  
 RA Fujimori Y., Komiyama M., Tashiro H., Tanigami A., Fujiwara T.,  
 RA Ono T., Yamada K., Fujii Y., Ozaki K., Hirao M., Ohmori Y.,  
 RA Kawabata A., Hikiji T., Kobatake N., Inagaki H., Ikema Y., Okamoto S.,  
 RA Okitani R., Kawakami T., Noguchi S., Itoh T., Shigeta K., Senba T.,  
 RA Matsumura K., Nakajima Y., Mizuno T., Morinaga M., Sasaki M.,  
 RA Togashi T., Oyama M., Hata H., Watanabe M., Komatsu T.,  
 RA Mizushima-Sugano J., Satoh T., Shirai Y., Takahashi Y., Nakagawa K.,  
 RA Okumura K., Nagase T., Nomura N., Kikuchi H., Masuho Y., Yamaashita R.,  
 RA Nakai K., Yada T., Nakamura Y., Ohara O., Isogai T., Sugano S.;  
 RT "Complete sequencing and characterization of 21,243 full-length human  
 cDNAs.";  
 RL Nat. Genet. 36:40-45(2004).  
 CC -I- FUNCTION: Involved in the biosynthesis of L2/HNK-1 carbohydrate  
 CC epitope on both glycolipids and glycoproteins (By similarity).  
 CC -I- CATALYTIC ACTIVITY: UDP-glucuronate + 3-beta-D-galactosyl-4-beta-  
 CC D-galactosyl-O-beta-D-xylosylprotein = UDP + 3-beta-D-  
 CC glucuronosyl-3-beta-D-galactosyl-4-beta-D-galactosyl-O-beta-D-  
 CC xylosylprotein.  
 CC -I- COFACTOR: Manganese.  
 CC -I- PATHWAY: Glycosylation.  
 CC -I- SUBUNIT: Homodimer (Potential).  
 CC -I- SUBCELLULAR LOCATION: Type II membrane protein. Golgi.  
 CC -I- TISSUE SPECIFICITY: Expressed in the trachea, retina, spinal cord,  
 CC hippocampus and other brain regions, and, at lower levels, in  
 CC testis and ovary.  
 CC -I- SIMILARITY: Belongs to the glycosyltransferase 43 family.  
 CC  
 CC This Swiss-Prot entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use as long as its content is in no way modified and this statement is not  
 CC removed.

DR EMBL: AY070019; AAL57718.1; -; mRNA.  
 DR EMBL: AY070110; AAL58977.1; -; Genomic DNA.  
 DR EMBL: AY070108; AAL58977.1; JOINED; Genomic DNA.  
 DR EMBL: AY070103; AAL58977.1; JOINED; Genomic DNA.  
 DR EMBL: AB075843; BAB85549.1; -; ALT INIT; mRNA.  
 DR EMBL: AL121961; CA142145.1; -; Genomic DNA.  
 DR EMBL: AL450320; CA142145.1; JOINED; Genomic DNA.  
 DR EMBL: AL450320; CA139582.1; -; Genomic DNA.  
 DR EMBL: AL121961; CA139582.1; JOINED; Genomic DNA.  
 DR EMBL: AK055248; BAB70889.1; ALT INIT; mRNA.  
 DR HSSP: O94766; 1KNS.  
 DR Ensemble: ENSG00000112309; Homo sapiens.  
 DR HGNC: HGNC:922; B3GAT2.  
 DR H-invDB: HIX0006000; -.  
 DR MIM: 607497; -.  
 DR GO: GO:0016021; C:integral to membrane; ISS.  
 DR GO: GO:0015018; P:galactosylgalactosylxylosylprotein 3-beta-g. .; ISS.  
 DR GO: GO:0016757; P:transferase activity, transferring glycosyl. .; ISS.  
 DR GO: GO:0016051; P:carbohydrate biosynthesis; ISS.  
 DR InterPro: IPR005027; Glyco\_trans\_43.  
 DR PANTHER: PTHR10896; Glyco\_trans\_43; 1.  
 DR Pfam: PF03360; Glyco\_transf\_43; 1.  
 KW Glycoprotein; Golgi stack; Manganese; Metal-binding; Multigene family;  
 KW Signal-anchor; Transferase; Transmembrane.  
 FT TOPO\_DOM 1 2 Cytolasmic (Potential).  
 FT TRANSMEM 3 23 Signal-anchor for type II membrane  
 FT TOPO\_DOM 24 323 Lumenal (Potential).  
 FT ACT\_SITE 273 273 Proton acceptor (By similarity).  
 FT METAL 187 187 Manganese (By similarity).  
 FT CARBOHYD 67 67 N-linked (GlcNAc...) (Potential).  
 FT CARBOHYD 292 292 N-linked (GlcNAc...) (Potential).  
 SQ SEQUENCE 323 AA; 36919 MW; 85058D52D2D28463 CRC64;  
 Query Match 100.0%; Score 323; DB 1; Length 323;  
 Best Local Similarity 100.0%; Pred. No. 9.6e-312;  
 Matches 323; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 MKSALFTRFFILLPWILVIMLDVDRPVPPLTPRPFYSPYAVGRGARLPLRRGGPA 60  
 DB 1 MKSALFTRFFILLPWILVIMLDVDRPVPPLTPRPFYSPYAVGRGARLPLRRGGPA 60  
 QY 61 HGTQKNSRQPOPEPQLPTIYATPTYSRVRQKELTRLANTPROVAQLHWILVEDAA 120  
 DB 61 HGTQKNSRQPOPEPQLPTIYATPTYSRVRQKELTRLANTPROVAQLHWILVEDAA 120  
 QY 121 ARSELVSRFLARAGLPSTHLHVPTPRRYKRPGLPRATEORNAGLAWLRHORHQAQPGV 180  
 DB 121 ARSELVSRFLARAGLPSTHLHVPTPRRYKRPGLPRATEORNAGLAWLRHORHQAQPGV 180  
 QY 181 LFFADDDNTYSLELFQEMRTTRKVSVPVGLVGGRRYERPLVGVGVGWYTGWRADRP 240  
 DB 181 LFFADDDNTYSLELFQEMRTTRKVSVPVGLVGGRRYERPLVGVGVGWYTGWRADRP 240  
 QY 241 AIDMAGFAVSLQVILSNPKAVFKRGSQGMQESFLQITTVVEELPKANNCTKVLVWH 300  
 DB 241 AIDMAGFAVSLQVILSNPKAVFKRGSQGMQESFLQITTVVEELPKANNCTKVLVWH 300  
 QY 301 TRTEKVNLANEPKYHLDTVKIEV 323  
 DB 301 TRTEKVNLANEPKYHLDTVKIEV 323  
 RESULT 2  
 Q599K0 PANTR PRELIMINARY; PRT; 323 AA.  
 AC Q599K0;  
 DT 10-MAY-2005 (TrEMBLrel. 30, Created)  
 DT 10-MAY-2005 (TrEMBLrel. 30, Last sequence update)  
 DT 10-MAY-2005 (TrEMBLrel. 30, Last annotation update)  
 DE Beta-3-glucuronyltransferase-S (EC 2.4.1.135).  
 GN Name-b3gat2;  
 OS Pan troglodytes (Chimpanzee).



QY	173	HORAQPGVLFADDDNTYSLFQEMRTTRTKVSWPVGLVGGRRYERPLVNGKVVGWYT	232
DB	176	HORAQPGVLFADDDNTYSLFQEMRTTRTKVSWPVGLVGGRRYERPLVNGKVVGWYT	235
QY	233	GWRADRPFAIDMAGFAVSLQVILSNPKAVFKRGSGPQMGSDFL	277
DB	236	GWRADRPFAIDMAGFAVSLQVILSNPKAVFKRRGSGPQMGSDFL	280

RESULT 5

QY	Q5CAS7	CHICK PRELIMINARY;	PRT;	304	AA.
AC	Q5CAS7				
DT	10-MAY-2005	(TEMBUREl. 30, Created)			
DT	10-MAY-2005	(TEMBUREl. 30, Last sequence update)			
DT	10-MAY-2005	(TEMBUREl. 30, Last annotation update)			
DE	Beta3-glucuronosyltransferase-S	(EC 2.4.1.135).			
GN	Name=B3GAT2;				
OS	Gallus gallus (Chicken).				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
OC	Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;				
OC	Gallus.				
OX	NCBI_TaxID=9031;				
RN	[1]				
RP	NUCLEOTIDE SEQUENCE.				
RA	Quzzine M., Magdalo J., Fournel-Gigleux S., Mollicone R., Oriol R.;				
RT	"Phylogeny of beta3-glucuronosyltransferases."				
RL	Submitted (MAR-2005) to the EMBL/GenBank/DBJ databases.				
DR	EMBL: AJ889606; CA163864.1; -; mRNA.				
DR	InterPro: IPR005027; Glyco_trans_43.				
DR	PANTHER: PTHR10896; Glyco_trans_43; 6.				
DR	Pfam: PF03360; Glyco_transf_43; 1.				
KW	Glycosyltransferase; Transferase.				
QY	SEQUENCE	304	AA;	34618	MW; 5C45DF677FBEA624

Query Match 19.2%; Score 62; DB 2; Length 304;  
Best Local Similarity 100.0%; Pred. No. 1.6e-52;  
Matches 62; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	258	PKAVFKRRGSPQMGSDFLQKQITVVEELPKANNCTKVLVHTRTEKVNLANEPKYHLD	317
DB	239	PKAVFKRRGSPQMGSDFLQKQITVVEELPKANNCTKVLVHTRTEKVNLANEPKYHLD	298
QY	318	TV 319	
DB	299	TV 300	

RESULT 6

ID	B3GA2	MOUSE	STANDARD;	PRT;	324	AA.
AC	P59270;					
DT	28-FEB-2003	(Rel. 41, Created)				
DT	28-FEB-2003	(Rel. 41, Last sequence update)				
DT	13-SEP-2005	(Rel. 48, Last annotation update)				
DE	Galactosylgalactosylxyloprotein 3-beta-glucuronosyltransferase 2					
DE	(EC 2.4.1.135) (Beta-1,3-glucuronosyltransferase 2)					
DE	(Glucuronosyltransferase-S) (GLCAT-S) (UDP-glucuronosyltransferase-S)					
DE	(GLCAT-D).					
GN	Name=B3gat2; Synonyms=Glcats;					
OS	Mus musculus (Mouse).					
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;					
OC	Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;					
OC	Muroidea; Muridae; Murinae; Mus.					
OX	NCBI_TaxID=10090;					
RN	[1]					
RP	NUCLEOTIDE SEQUENCE (ISOFORM 1).					
RC	STRAIN=BALB/c; TISSUE=Brain;					
RX	MEDLINE=22271538; PubMed=12383500; DOI=10.1016/S0378-1119(02)00840-5;					
RA	Imiya K., Iehizaki T., Seiki T., Saito F., Inazawa J., Oka S.,					
RA	Kawasaki T.;					
RT	"cDNA cloning, genomic structure and chromosomal mapping of the mouse					

RT	glucuronosyltransferase-S involved in the biosynthesis of the HNK-1
RT	carbohydrate epitope.";
RL	Gene 296:29-36(2002).
RN	[2]
RP	NUCLEOTIDE SEQUENCE [LARGE SCALE MRNA] (ISOFORMS 1 AND 2).
RC	STRAIN=C57BL/6J; TISSUE=Embryonic head, and Kidney;
RX	MEDLINE=22354683; PubMed=12466851; DOI=10.1038/nature01266;
RA	Okazaki Y., Furuno M., Kasukawa T., Adachi J., Bono H., Kondo S.,
RA	Nakaido I., Oosato N., Saito R., Suzuki H., Yamanaka I., Kiyosawa H.,
RA	Yagi K., Tomaru Y., Hasegawa Y., Nogami A., Schonbach C., Gojobori T.,
RA	Baldarelli R., Hill D.P., Bult C., Hume D.A., Quackenbush J.,
RA	Schriml L.M., Kanapin A., Matsuda H., Batalov S., Beisel K.W.,
RA	Blake J.A., Bradt D., Brusic V., Chothia C., Corbani L.E., Cousins S.,
RA	Dalla E., Dragani T.A., Fletcher C.F., Forrest A., Frazer K.S.,
RA	Gaasterland T., Gariboldi M., Gissi C., Godzik A., Gough J.,
RA	Grimmond S., Gustincich S., Hirokawa N., Jackson I.J., Jarvis E.D.,
RA	Kanai A., Kawaji H., Kawasawa Y., Kedzierski R.M., King B.L.,
RA	Konagaya A., Kurochkin I.V., Lee Y., Lenhard B., Lyons P.A.,
RA	Maglott D.R., Maltais L., Marchionni L., McKenzie L., Miki H.,
RA	Nagashima T., Numata K., Okido T., Pavan W.J., Pertea G., Pesole G.,
RA	Petrovsky N., Pillai R., Pontius J.U., Qi D., Ramachandran S.,
RA	Ravasi T., Reed J.C., Reid D.J., Reid J., Ring B.Z., Ringwald M.,
RA	Sandelin A., Schneider C., Semple C.A., Setou M., Shimada K.,
RA	Sultana R., Takenaka Y., Taylor M.S., Teasdale R.D., Tomita M.,
RA	Verardo R., Wagner L., Wahlestedt C., Wang Y., Watanabe Y., Wells C.,
RA	Wilming L.G., Wynshaw-Boris A., Yanagisawa M., Yang I., Yang L.,
RA	Yuan Z., Zvolan M., Zhu Y., Zimmer A., Carninci P., Hayatsu N.,
RA	Hirozane-Kishikawa T., Konno H., Nakamura M., Sakazume N., Sato K.,
RA	Shiraki T., Waki K., Kawai J., Aizawa K., Arakawa T., Fukuda S.,
RA	Hara A., Hashizume W., Imotani K., Iehii Y., Itoh M., Kagawa I.,
RA	Miyazaki A., Sakai K., Sasaki D., Shibata K., Shinagawa A.,
RA	Yasunishi A., Yoshino M., Waterston R., Lander E.S., Rogers J.,
RA	Birney E., Hayashizaki Y.;
RT	"Analysis of the mouse transcriptome based on functional annotation of
RT	60,770 full-length cDNAs.";
RL	Nature 420:563-573(2002).
RN	[3]
RP	NUCLEOTIDE SEQUENCE [LARGE SCALE MRNA] (ISOFORM 1).
RC	STRAIN=C57BL/6; TISSUE=Brain;
RX	MEDLINE=2238257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA	Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA	Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA	Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA	Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA	Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA	Stapleton M., Soares M.B., Bonaldo M.P., Casavant T.L., Scheetz T.E.,
RA	Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA	Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA	Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA	Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA	Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA	Fahey J., Helton E., Kettman M., Madan A., Rodrigues S., Sanchez A.,
RA	Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA	Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA	Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA	Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smallos D.E.,
RA	Schmerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT	"Generation and initial analysis of more than 15,000 full-length human
RT	and mouse cDNA sequences.";
RL	Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
CC	-!- FUNCTION: Involved in the biosynthesis of L2/HNK-1 carbohydrate
CC	epitope on both glycolipids and glycoproteins.
CC	-!- CATALYTIC ACTIVITY: UDP-glucuronate + 3-beta-D-galactosyl-4-beta-
CC	D-galactosyl-O-beta-D-xylosylprotein + 3-beta-D-
CC	glucuronosyl-3-beta-D-galactosyl-4-beta-D-galactosyl-O-beta-D-
CC	xylosylprotein.
CC	-!- COFACTOR: Manganese.
CC	-!- PATHWAY: Glycosylation.
CC	-!- SUBUNIT: Homodimer (potential).
CC	-!- SUBCELLULAR LOCATION: Type II membrane protein. Golgi (By
CC	similarity).
CC	-!- ALTERNATIVE PRODUCTS:
CC	Event-Alternative splicing; Named isoforms=2;

```
CC Name=1;
CC IsoId=P59270-1; Sequence=Displayed;
CC Name=2;
CC IsoId=P59270-2; Sequence=VSP_001796, VSP_001797;
CC Note=No experimental confirmation available;
CC -1- TISSUE SPECIFICITY: Expressed in brain, but not in liver and
CC kidney.
CC -1- SIMILARITY: Belongs to the glycosyltransferase 43 family.
CC -----
CC This Swiss-Prot entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use as long as its content is in no way modified and this statement is not
CC removed.
CC -----
CC EMBL: AB055902; BAC20343.1; -; mRNA.
CC EMBL: AK048146; BAC33257.1; -; mRNA.
CC EMBL: AK052640; BAC35075.1; -; mRNA.
CC EMBL: BC056368; AAH56368.1; -; mRNA.
CC EMBL: BC058082; AAH58082.1; -; mRNA.
CC HSSP: O94766; 1KWS.
CC Ensembl: ENSMUSG0000026156; Mus musculus.
CC MGI: MGI:2389490; B3gat2.
CC GO: GO:0016021; C: integral to membrane; ISS.
CC GO: GO:0015018; F: galactosylgalactosylxylosylprotein 3-beta-g. .; ISS.
CC GO: GO:0016757; F: transferase activity, transferring glycosyl. .; ISS.
CC GO: GO:0016051; P: carbohydrate biosynthesis; ISS.
CC InterPro: IPR005027; Glyco_trans_43.
CC PANTHER: PTHR10896; Glyco_trans_43; 1.
CC Pfam: PF03360; Glyco_transf_43; 1.
CC Alternative splicing: Glycoprotein; Golgi stack; Manganese;
CC Metal-binding; Multigene family; Signal-anchor; Transferase;
CC Transmembrane.
CC TOPO_DOM 1 2 Cytoplasmic (Potential).
CC TRANSMEM 3 23 Signal-anchor for type II membrane
CC protein (Potential).
CC TOPO_DOM 24 324 Lumenal (Potential).
CC ACT_SITE 274 274 Proton acceptor (By similarity).
CC METAL 188 188 Manganese (By similarity).
CC CARBOHYD 68 68 N-linked (GlcNAc. . .) (Potential).
CC CARBOHYD 293 293 N-linked (GlcNAc. . .) (Potential).
CC VARSPPLIC 248 302 FAVSLQVILSNPKAVFKRRGSGPQMGSDFLKQITVVELEPKA
CC PKASNCTKVLVWHT -> EQNWDICPCRMGRNWRPTF
CC IKPLSVLHSSEILLIPKIGSLWQVEVDFRSG (in
CC isoform 2).
CC /FTid=VSP_001796.
CC Missing (in isoform 2).
CC VARSPPLIC 303 324 /FTid=VSP_001797.
CC CONFLICT 291 291 A -> V (in Ref. 1).
CC SEQUENCE 324 AA; 37132 MW; 675C549D3787B957 CRC64;
CC
CC Query Match 16.7%; Score 54; DB 1; Length 324;
CC Best Local Similarity 100.0%; Pred. No. 1.5e-44;
CC Matches 54; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
CC
CC QY 237 DRPFAIDMAGFAVSLQVILSNPKAVFKRRGSGPQMGSDFLKQITVVELEPKA 290
CC
CC Db 238 DRPFAIDMAGFAVSLQVILSNPKAVFKRRGSGPQMGSDFLKQITVVELEPKA 291
CC
CC RESULT 7
CC QSDTT7 MOUSE PRELIMINARY; PRT; 349 AA.
CC AC QSDTT7;
CC DT 10-MAY-2005 (TrEMBLrel. 30, Created)
CC DT 10-MAY-2005 (TrEMBLrel. 30, Last sequence update)
CC DT 10-MAY-2005 (TrEMBLrel. 30, Last annotation update)
CC DE MKIAA1963 protein (Fragment).
CC GN Name=B3gat2; Synonyms=MKIAA1963;
CC OS Mus musculus (Mouse).
CC OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
CC OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
CC OC Muridae; Murinae; Mus.
```

```
OX NCBI_TaxID=10090;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC TISSUE=Retal brain;
RA Okazaki N., Kikuno R.F., Ohara R., Inamoto S., Nagase T., Ohara O.,
RA Koga H.;
RT "Prediction of the Coding Sequences of Mouse Homologues of KIAA Gene.
RT The Complete Nucleotide Sequences of Mouse KIAA-homologous cDNAs
RT Identified by Screening of Terminal sequences of cDNA Clones Randomly
RT Sampled from Size-Fractionated Libraries. ";
RL Submitted (FEB-2005) to the EMBL/GenBank/DBJ databases.
DR EMBL: AK220433; BAD90265.1; -; mRNA.
DR MGI: MGI:2389490; B3gat2.
DR GO: GO:0015020; F: glucuronosyltransferase activity; TAS.
DR InterPro: IPR005027; Glyco_transf_43.
DR Pfam: PF03360; Glyco_transf_43; 1.
FT NON_TER 1
SQ SEQUENCE 349 AA; 40066 MW; 2EPB2B8BCBD84169 CRC64;
CC
CC Query Match 16.7%; Score 54; DB 2; Length 349;
CC Best Local Similarity 100.0%; Pred. No. 1.6e-44;
CC Matches 54; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
CC
CC QY 237 DRPFAIDMAGFAVSLQVILSNPKAVFKRRGSGPQMGSDFLKQITVVELEPKA 290
CC
CC Db 263 DRPFAIDMAGFAVSLQVILSNPKAVFKRRGSGPQMGSDFLKQITVVELEPKA 316
CC
CC RESULT 8
CC B3GA2 RAT STANDARD; PRT; 324 AA.
CC ID B3GA2 RAT
CC AC Q92137;
CC DT 16-OCT-2001 (Rel. 40, Created)
CC DT 15-OCT-2001 (Rel. 40, Last sequence update)
CC DT 13-SEP-2005 (Rel. 48, Last annotation update)
CC DE Galactosylgalactosylxylosylprotein 3-beta-glucuronosyltransferase 2
CC DE (BC 2.4.1.135) (Beta-1,3-glucuronyltransferase 2)
CC DE (Glucuronosyltransferase-S) (GlcAT-S) (UDP-glucuronosyltransferase-S)
CC DE (GlcAT-D).
CC GN Name=B3gat2; Synonyms=Glcats;
CC OS Rattus norvegicus (Rat).
CC OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
CC OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
CC OC Muroidae; Muridae; Murinae; Rattus.
CC OX NCBI_TaxID=10116;
CC RN [1]
CC RP NUCLEOTIDE SEQUENCE.
CC RC TISSUE=Brain;
CC RX MEDLINE=99185317; PubMed=10082676; DOI=10.1006/bbrc.1999.0151;
CC RA Seiki T., Oka S., Terayama K., Imiya K., Kawasaki T.;
CC RT "Molecular cloning and expression of a second glucuronyltransferase
CC involved in the biosynthesis of the HNK-1 carbohydrate epitope. ";
CC RL Biochem. Biophys. Res. Commun. 255:182-187(1999).
CC RN [2]
CC RP NUCLEOTIDE SEQUENCE, AND CHARACTERIZATION.
CC RC TISSUE=Embryonic brain;
CC RX MEDLINE=99287914; PubMed=10358066; DOI=10.1074/jbc.274.24.17115;
CC RA Shimoda Y., Tajima Y., Nagase T., Harii K., Osumi N., Sanai Y.;
CC RT "Cloning and expression of a novel galactoside beta1, 3-
CC glucuronyltransferase involved in the biosynthesis of HNK-1 epitope. ";
CC RL J. Biol. Chem. 274:17115-17122(1999).
CC CC -1- FUNCTION: Involved in the biosynthesis of L2/HNK-1 carbohydrate
CC epitope on both glycolipids and glycoproteins. Substrates include
CC asialo-orosomucoid (ASOR), paragloboside (lacto-N-
CC neotetraosylceramide), Gal-beta-1,4-GlcNAc-beta-1,3-Gal-beta-1,4-
CC Glc-pyridylamine and Gal-beta-1,3-GlcNAc-beta-1,3-Gal-beta-1,4-
CC Glc-pyridylamine.
CC CC -1- CATALYTIC ACTIVITY: UDP-glucuronate + 3-beta-D-galactosyl-4-beta-
CC D-galactosyl-O-beta-D-xylosylprotein = UDP + 3-beta-D-
CC glucuronosyl-3-beta-D-galactosyl-4-beta-D-galactosyl-O-beta-D-
CC xylosylprotein.
CC CC -1- COFACTOR: Manganese.
CC CC -1- PATHWAY: Glycosylation.
```





```
RESULT 12
Q599J9_ORYLA
ID Q599J9_ORYLA PRELIMINARY; PRT; 303 AA.
AC Q599J9;
DT 10-MAY-2005 (TrEMBLrel. 30, Created)
DT 10-MAY-2005 (TrEMBLrel. 30, Last sequence update)
DT 10-MAY-2005 (TrEMBLrel. 30, Last annotation update)
DE Beta-3-glucuronyltransferase-S (EC 2.4.1.135).
GN Name=b3gat2;
OS Oryzias latipes (Medaka fish) (Japanese ricefish).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorpha; Acanthopterygii; Percomorpha; Atherinomorpha;
OC Belontiiformes; Adrianichthyidae; Oryziinae; Oryzias.
OX NCBI_TaxID=8090;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RA Ouzzine M., Magdalou J., Fournel-Gigleux S., Mollicone R., Oriol R.;
RT "Phylogeny of beta3-glucuronyltransferases.";
RL Submitted (MAR-2005) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ890842; CA168028.1; -; mRNA.
DR INTERPRO; IPR005027; TPR-like_helical.
DR PANTHER; PTHR10896; Glyco_trans_43; 7.
DR Pfam; PF03360; Glyco_transf_43; 1.
KW Glycosyltransferase; Transferase.
SQ SEQUENCE 303 AA; 33932 MW; 6CD039E7361589A7 CRC64;

Query Match 6.8%; Score 22; DB 2; Length 303;
Best Local Similarity 100.0%; Pred. No. 8.4e-13;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 84 AITPTYSRPVQKAELTRLANTF 105
DQ 67 AITPTYSRPVQKAELTRLANTF 88

RESULT 13
Q5CB02_BRARE
ID Q5CB02_BRARE PRELIMINARY; PRT; 334 AA.
AC Q5CB02;
DT 10-MAY-2005 (TrEMBLrel. 30, Created)
DT 10-MAY-2005 (TrEMBLrel. 30, Last sequence update)
DT 10-MAY-2005 (TrEMBLrel. 30, Last annotation update)
DE Beta3-glucuronyltransferase (EC 2.4.1.135).
GN Name=b3gat1;
OS Brachydanio rerio (Zebrafish) (Danio rerio).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC Cyprinidae; Danio.
OX NCBI_TaxID=7955;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RA Ouzzine M., Magdalou J., Fournel-Gigleux S., Mollicone R., Oriol R.;
RT "Phylogeny of beta3-glucuronyltransferases.";
RL Submitted (MAR-2005) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ888976; CA162041.1; -; mRNA.
DR InterPro; IPR005027; Glyco_trans_43.
DR InterPro; IPR011990; TPR-like_helical.
DR PANTHER; PTHR10896; Glyco_trans_43; 7.
DR Pfam; PF03360; Glyco_transf_43; 1.
KW Glycosyltransferase; Transferase.
SQ SEQUENCE 334 AA; 38370 MW; 4F5556173941A4C2 CRC64;

Query Match 6.5%; Score 21; DB 2; Length 334;
Best Local Similarity 100.0%; Pred. No. 8.9e-12;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 85 ITPTYSRPVQKAELTRLANTF 105
DQ 89 ITPTYSRPVQKAELTRLANTF 109

RESULT 14
Q5CB01_ORYLA
ID Q5CB01_ORYLA PRELIMINARY; PRT; 335 AA.
```

```
Q5CB01;
DT 10-MAY-2005 (TrEMBLrel. 30, Created)
DT 10-MAY-2005 (TrEMBLrel. 30, Last sequence update)
DT 10-MAY-2005 (TrEMBLrel. 30, Last annotation update)
DE Beta3-glucuronyltransferase (EC 2.4.1.135).
GN Name=b3gat1;
OS Oryzias latipes (Medaka fish) (Japanese ricefish).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorpha; Acanthopterygii; Percomorpha; Atherinomorpha;
OC Belontiiformes; Adrianichthyidae; Oryziinae; Oryzias.
OX NCBI_TaxID=8090;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RA Ouzzine M., Magdalou J., Fournel-Gigleux S., Mollicone R., Oriol R.;
RT "Phylogeny of beta3-glucuronyltransferases.";
RL Submitted (MAR-2005) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ888977; CA162042.1; -; mRNA.
DR InterPro; IPR005027; Glyco_trans_43.
DR InterPro; IPR011990; TPR-like_helical.
DR PANTHER; PTHR10896; Glyco_trans_43; 7.
DR Pfam; PF03360; Glyco_transf_43; 1.
KW Glycosyltransferase; Transferase.
SQ SEQUENCE 335 AA; 38244 MW; 365B1E0E76C10B31 CRC64;

Query Match 6.5%; Score 21; DB 2; Length 335;
Best Local Similarity 100.0%; Pred. No. 9e-12;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 85 ITPTYSRPVQKAELTRLANTF 105
DQ 90 ITPTYSRPVQKAELTRLANTF 110

RESULT 15
Q5CAZ9_TETNG
ID Q5CAZ9_TETNG PRELIMINARY; PRT; 335 AA.
AC Q5CAZ9;
DT 10-MAY-2005 (TrEMBLrel. 30, Created)
DT 10-MAY-2005 (TrEMBLrel. 30, Last sequence update)
DT 10-MAY-2005 (TrEMBLrel. 30, Last annotation update)
DE Beta3-glucuronyltransferase (EC 2.4.1.135).
GN Name=b3gat1;
OS Tetraodon nigroviridis (Green puffer).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
OC Tetraodontidae; Tetraodontidae; Tetraodon.
OX NCBI_TaxID=99883;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RA Ouzzine M., Magdalou J., Fournel-Gigleux S., Mollicone R., Oriol R.;
RT "Phylogeny of beta3-glucuronyltransferases.";
RL Submitted (MAR-2005) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ888979; CA162044.1; -; mRNA.
DR InterPro; IPR005027; Glyco_trans_43.
DR InterPro; IPR011990; TPR-like_helical.
DR PANTHER; PTHR10896; Glyco_trans_43; 7.
DR Pfam; PF03360; Glyco_transf_43; 1.
KW Glycosyltransferase; Transferase.
SQ SEQUENCE 335 AA; 38292 MW; F89E348F901C5A25 CRC64;

Query Match 6.5%; Score 21; DB 2; Length 335;
Best Local Similarity 100.0%; Pred. No. 9e-12;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 85 ITPTYSRPVQKAELTRLANTF 105
DQ 90 ITPTYSRPVQKAELTRLANTF 110

Search completed: May 9, 2006, 10:58:11
```

Job time : 60 secs

---

GenCore version 5.1.8  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: May 9, 2006, 10:57:17 ; Search time 27 Seconds  
(without alignments)  
989.046 Million cell

**Title:** US-10-634-905-2

Perfect score:

Sequence: 1 MKSALFTRFFILLPWILIVI.....EKNVLANEPKYHLDTVKIEV 323

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 572060 segs. 82675679 residues

Word size : 1

Total number of hits satisfying chosen parameters: 570988

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

### Post-processing: Listing first 45 summaries

Database : Issued Patents AA:\*

```
1: /cgn2 6/ptodata/1/iaa/5 COMB.pcp:*
```

2: /cgn2\_6/ptodata/1/iaa/6\_COMB.pep.\*

3: /cgn2\_6/ptodata/1/iaa/H\_COMB.pcp:\*

4: /cgn2\_6/ptodata/1/iaa/PCTUS\_COMB.pep:\*

5: /cgn2\_6/ptodata/1/iaa/RE COMB.pcp:\*

6: /cgn2\_6/ptodata/1/iaa/backfiles1.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	323	100.0	323	2	US-09-816-095-2	Sequence 2, Appli
2	197	61.0	197	2	US-09-816-095-5	Sequence 5, Appli
3	47	14.6	324	2	US-09-816-095-4	Sequence 4, Appli
4	15	4.6	347	2	US-09-059-369-2	Sequence 2, Appli
5	11	3.4	22	2	US-09-059-369-5	Sequence 5, Appli
6	10	3.1	63	2	US-09-621-976-7125	Sequence 7125, Ap
7	8	2.5	25	2	US-09-059-369-4	Sequence 4, Appli
8	8	2.5	107	2	US-09-328-352-5562	Sequence 5562, Ap
9	8	2.5	142	2	US-09-270-767-45263	Sequence 45263, A
10	8	2.5	256	2	US-09-252-991A-25670	Sequence 25670, A
11	8	2.5	405	2	US-09-252-991A-16887	Sequence 16887, A
12	8	2.5	410	2	US-09-252-991A-30530	Sequence 30570, A
13	8	2.5	530	2	US-09-252-991A-23339	Sequence 23339, A
14	8	2.5	2628	2	US-09-413-814-11	Sequence 11, Appl
15	7	2.2	80	2	US-09-621-976-6610	Sequence 6610, Ap
16	7	2.2	90	2	US-09-949-002-438	Sequence 438, App
17	7	2.2	117	2	US-09-270-767-41066	Sequence 41066, A
18	7	2.2	117	2	US-09-270-767-56282	Sequence 56282, A
19	7	2.2	131	2	US-09-902-540-11259	Sequence 11259, A
20	7	2.2	135	2	US-09-489-039A-10174	Sequence 10174, A
21	7	2.2	172	2	US-09-252-991A-25750	Sequence 25750, A
22	7	2.2	181	2	US-09-605-703B-1344	Sequence 1344, Ap
23	7	2.2	208	2	US-09-252-991A-27197	Sequence 27197, A
24	7	2.2	237	1	US-08-103-998-4	Sequence 4, Appli
25	7	2.2	249	2	US-09-328-352-5813	Sequence 5813, Ap
26	7	2.2	254	2	US-09-902-540-16721	Sequence 16721, A
27	7	2.2	257	2	US-09-252-991A-21740	Sequence 21740, A

```
US-09-816-095-5
; Sequence 5, Application US/09816095
; Patent No. 6664084
; GENERAL INFORMATION:
; APPLICANT: GAN, Weiniu
; TITLE OF INVENTION: ISOLATED HUMAN ENZYME PROTEINS, NUCLEIC
; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN ENZYME PROTEINS, AND USES
; TITLE OF INVENTION: THEREOF
; FILE REFERENCE: CL001147
; CURRENT APPLICATION NUMBER: US/09/816,095
; CURRENT FILING DATE: 2001-03-26
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5
; LENGTH: 197
; TYPE: PRT
; ORGANISM: Human
US-09-816-095-5
Query Match          61.0%; Score 197; DB 2; Length 197;
Best Local Similarity 100.0%; Pred. No. 5.3e-190;
Matches 197; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MKSALTRFILLPWILVIMLDVTRRPVPLTPRPYFSPYAVGGRGRLPLRRGGA 60
Db 1 MKSALTRFILLPWILVIMLDVTRRPVPLTPRPYFSPYAVGGRGRLPLRRGGA 60
QY 61 HGQTKNQSRQPQPEPQLPTIYAITPTYSRPVQKAELTRLANTFPQVQLHWILVEDAA 120
Db 61 HGQTKNQSRQPQPEPQLPTIYAITPTYSRPVQKAELTRLANTFPQVQLHWILVEDAA 120
QY 121 ARSELVSRFLARAGLSTHLHVTPTRYPKPGLPRAEQBNAGLAWLRQHQHQAQPGV 180
Db 121 ARSELVSRFLARAGLSTHLHVTPTRYPKPGLPRAEQBNAGLAWLRQHQHQAQPGV 180
QY 181 LFFADDDNTYSLFQEB 197
Db 181 LFFADDDNTYSLFQEB 197
RESULT 3
US-09-816-095-4
; Sequence 4, Application US/09816095
; Patent No. 6664084
; GENERAL INFORMATION:
; APPLICANT: GAN, Weiniu
; TITLE OF INVENTION: ISOLATED HUMAN ENZYME PROTEINS, NUCLEIC
; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN ENZYME PROTEINS, AND USES
; TITLE OF INVENTION: THEREOF
; FILE REFERENCE: CL001147
; CURRENT APPLICATION NUMBER: US/09/816,095
; CURRENT FILING DATE: 2001-03-26
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 324
; TYPE: PRT
; ORGANISM: Rattus norvegicus
US-09-816-095-4
Query Match          14.6%; Score 47; DB 2; Length 324;
Best Local Similarity 100.0%; Pred. No. 7e-39;
Matches 47; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 176 AQFGVLFFADDDNTYSLFQEMRTTRKYSVWPVGLVGGRYRPLV 222
Db 177 AQFGVLFFADDDNTYSLFQEMRTTRKYSVWPVGLVGGRYRPLV 223
RESULT 4
US-09-059-369-2
; Sequence 2, Application US/09059369
; Patent No. 6040156
; GENERAL INFORMATION:
; APPLICANT: KAWASAKI, TOSHISUKE
; TITLE OF INVENTION: DNA ENCODING GLUCURONYLTRANSFERASE
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,
; STREET: 1755 S. JEFFERSON DAVIS HIGHWAY, FOURTH FLOOR
; CITY: ARLINGTON
; STATE: VA
; COUNTRY: USA
; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/059,369
; FILING DATE: 14-APR-1998
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 9-127065
; FILING DATE: 16-MAY-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: OBLON, NORMAN F.
; REGISTRATION NUMBER: 24,618
; REFERENCE/DOCKET NUMBER: 9378-0002-0
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703-413-3000
; TELEFAX: 703-413-2220
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 347 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-059-369-2
Query Match          4.6%; Score 15; DB 2; Length 347;
Best Local Similarity 100.0%; Pred. No. 1.2e-06;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 86 TPTYSRPVQKAELTR 100
Db 103 TPTYSRPVQKAELTR 117
RESULT 5
US-09-059-369-5
; Sequence 5, Application US/09059369
; Patent No. 6040156
; GENERAL INFORMATION:
; APPLICANT: KAWASAKI, TOSHISUKE
; TITLE OF INVENTION: DNA ENCODING GLUCURONYLTRANSFERASE
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,
; STREET: 1755 S. JEFFERSON DAVIS HIGHWAY, FOURTH FLOOR
; CITY: ARLINGTON
; STATE: VA
; COUNTRY: USA
; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/059,369
```

; FILING DATE: 14-APR-1998  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: JP 9-127065  
; FILING DATE: 16-MAY-1997  
; ATTORNEY/AGENT INFORMATION:  
; NAME: OBLON, NORMAN F.  
; REGISTRATION NUMBER: 24,618  
; REFERENCE/DOCKET NUMBER: 9378-0002-0  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 703-413-3000  
; TELEFAX: 703-413-2220  
; INFORMATION FOR SEQ ID NO: 5:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 22 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
US-09-059-369-5

Query Match 3.4%; Score 11; DB 2; Length 22;  
Best Local Similarity 100.0%; Pred. No. 0.00097;  
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 239 PFAIDMAGFAV 249  
Db 8 PFAIDMAGFAV 18  
|||||

## RESULT 6

US-09-621-976-7125  
; Sequence 7125, Application US/09621976  
; Patent No. 6639063  
; GENERAL INFORMATION:  
; APPLICANT: Dumas Milne Edwards, J.B.  
; APPLICANT: Jobert, S.  
; APPLICANT: Giordano, J.Y.  
; TITLE OF INVENTION: ESTs and Encoded Human Proteins.  
; FILE REFERENCE: GENSET.054PR2  
; CURRENT APPLICATION NUMBER: US/09/621.976  
; CURRENT FILING DATE: 2000-07-21  
; NUMBER OF SEQ ID NOS: 19335  
; SOFTWARE: Patent.pm  
; SEQ ID NO 7125  
; LENGTH: 63  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-621-976-7125

Query Match 3.1%; Score 10; DB 2; Length 63;  
Best Local Similarity 100.0%; Pred. No. 0.026;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 296 VLVVHTRTEK 305  
Db 32 VLVVHTRTEK 41  
|||||

## RESULT 7

US-09-059-369-4  
; Sequence 4, Application US/09059369  
; Patent No. 6040156  
; GENERAL INFORMATION:  
; APPLICANT: KAWASAKI, TOSHISUKE  
; APPLICANT: OKA, SHOGO  
; TITLE OF INVENTION: DNA ENCODING GLUCURONYLTRANSFERASE  
; NUMBER OF SEQUENCES: 20  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: OBLON, SPIVAK, MCLELLAND, MAIER & NEUSTADT,  
; ADDRESS: P.C.  
; STREET: 1755 S. JEFFERSON DAVIS HIGHWAY, FOURTH FLOOR  
; CITY: ARLINGTON

; STATE: VA  
; COUNTRY: USA  
; ZIP: 22202  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/059,369  
; FILING DATE: 14-APR-1998  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: JP 9-127065  
; FILING DATE: 16-MAY-1997  
; ATTORNEY/AGENT INFORMATION:  
; NAME: OBLON, NORMAN F.  
; REGISTRATION NUMBER: 24,618  
; REFERENCE/DOCKET NUMBER: 9378-0002-0  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 703-413-3000  
; TELEFAX: 703-413-2220  
; INFORMATION FOR SEQ ID NO: 4:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 25 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
US-09-059-369-4

Query Match 2.5%; Score 8; DB 2; Length 25;  
Best Local Similarity 100.0%; Pred. No. 1.1;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 183 FADDDNTY 190  
Db 9 FADDDNTY 16  
|||||

## RESULT 8

US-09-328-352-5562  
; Sequence 5562, Application US/09328352  
; Patent No. 6562958  
; GENERAL INFORMATION:  
; APPLICANT: Gary L. Breton et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER  
; FILE REFERENCE: GTC99-03PA  
; CURRENT APPLICATION NUMBER: US/09/328,352  
; CURRENT FILING DATE: 1999-06-04  
; NUMBER OF SEQ ID NOS: 8252  
; SEQ ID NO 5562  
; LENGTH: 107  
; TYPE: PRT  
; ORGANISM: Acinetobacter baumannii  
US-09-328-352-5562

Query Match 2.5%; Score 8; DB 2; Length 107;  
Best Local Similarity 100.0%; Pred. No. 4.4;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 96 AELTRLAN 103  
Db 71 AELTRLAN 78  
|||||

## RESULT 9

US-09-270-767-45263  
; Sequence 45263, Application US/09270767  
; Patent No. 6703491  
; GENERAL INFORMATION:  
; APPLICANT: Homburger et al.

; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster

; FILE REFERENCE: File Reference: 7326-094  
; CURRENT APPLICATION NUMBER: US/09/270,767  
; CURRENT FILING DATE: 1999-03-17  
; NUMBER OF SEQ ID NOS: 62517  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 45263  
; LENGTH: 142  
; TYPE: PRT  
; ORGANISM: Drosophila melanogaster  
US-09-270-767-45263

Query Match 2.5%; Score 8; DB 2; Length 142;  
Best Local Similarity 100.0%; Pred. No. 5.6;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 243 DMAGFAVS 250  
Db 54 DMAGFAVS 61  
|||||

## RESULT 10

US-09-252-991A-25670  
; Sequence 25670, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 25670  
; LENGTH: 256  
; TYPE: PRT  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-25670

Query Match 2.5%; Score 8; DB 2; Length 256;  
Best Local Similarity 100.0%; Pred. No. 9.7;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 46 GGGARLP 53  
Db 24 GGGARLP 31  
|||||

## RESULT 11

US-09-252-991A-16887  
; Sequence 16887, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 16887  
; LENGTH: 405  
; TYPE: PRT  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-16887

Query Match 2.5%; Score 8; DB 2; Length 405;  
Best Local Similarity 100.0%; Pred. No. 15;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 173 HORAOPGV 180  
Db 270 HORAOPGV 277  
|||||

## RESULT 12

US-09-252-991A-30570  
; Sequence 30570, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 30570  
; LENGTH: 410  
; TYPE: PRT  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-30570

Query Match 2.5%; Score 8; DB 2; Length 410;  
Best Local Similarity 100.0%; Pred. No. 15;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 169 QRHQHQA 176  
Db 394 QRHQHQA 401  
|||||

## RESULT 13

US-09-252-991A-23339  
; Sequence 23339, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 23339  
; LENGTH: 530  
; TYPE: PRT  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-23339

Query Match 2.5%; Score 8; DB 2; Length 530;  
Best Local Similarity 100.0%; Pred. No. 19;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 209 VGLVGRR 216  
Db 463 VGLVGRR 470  
|||||

## RESULT 14

```
US-09-413-814-11
; Sequence 11, Application US/09413814
; Patent No. 6225064
; GENERAL INFORMATION:
; APPLICANT: Gesellschaft fuer Biotechnologische Forschung mbH
; APPLICANT: Bristol-Myers Squibb, Co.
; APPLICANT: Beyer, Stefan
; APPLICANT: Bloecker, Helmut
; APPLICANT: Brandt, Petra
; APPLICANT: Cino, Paul M
; APPLICANT: Dougherty, Brian A
; APPLICANT: Goldberg, Steven L
; APPLICANT: Hofle, Gerhard
; APPLICANT: Mueller, Joachim
; APPLICANT: Reichenbach, Hans
; TITLE OF INVENTION: DNA sequences for enzymatic synthesis of polyketide or
; TITLE OF INVENTION: heteropolysaccharide compounds
; FILE REFERENCE: PCT/US 99/23535
; CURRENT APPLICATION NUMBER: US/09/413,814
; CURRENT FILING DATE: 1999-10-07
; EARLIER APPLICATION NUMBER: DE 198 46 493.2
; EARLIER FILING DATE: 1998-10-09
; NUMBER OF SEQ ID NOS: 107
; SOFTWARE: Patentin ver. 2.1
; SEQ ID NO 11
; LENGTH: 2628
; TYPE: PRT
; ORGANISM: Sorangium cellulosum
US-09-413-814-11

Query Match      2.5%; Score 8; DB 2; Length 2628;
Best Local Similarity 100.0%; Pred. No. 83;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      45  VGRGGARL 52
Db      2456 VGRGGARL 2463

RESULT 15
US-09-621-976-6610
; Sequence 6610, Application US/09621976
; Patent No. 6639063
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Jobert, S.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: ESTs and Encoded Human Proteins.
; FILE REFERENCE: GENSET.054PR2
; CURRENT APPLICATION NUMBER: US/09/621,976
; CURRENT FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 19335
; SOFTWARE: Patent.pm
; SEQ ID NO 6610
; LENGTH: 80
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-621-976-6610

Query Match      2.2%; Score 7; DB 2; Length 80;
Best Local Similarity 100.0%; Pred. No. 34;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      274  SDFLKQI 280
Db      9     SDFLKQI 15

Search completed: May 9, 2006, 10:58:44
Job time : 27 secs
```

**This Page Blank (uspto)**



GenCore version 5.1.8  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: May 9, 2006, 10:58:21 ; Search time 78 Seconds  
(without alignments)  
1730.242 Million cell updates/sec

Title: US-10-634-905-2  
Perfect score: 323  
Sequence: 1 MKSALTRFFILLPWILVI.....EKVNLANEKPYHLDTVKIEV 323

Scoring table: OLIGO  
Gapop 60.0 , Gapext 60.0

Searched: 1867569 seqs, 417829326 residues

Word size : 1  
Total number of hits satisfying chosen parameters: 1866650

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Published Applications AA\_Main:\*  
1: /cgn2\_6/prodata/1/pubpaa/US07\_PUBCOMB.pep:\*  
2: /cgn2\_6/prodata/1/pubpaa/US08\_PUBCOMB.pep:\*  
3: /cgn2\_6/prodata/1/pubpaa/US09\_PUBCOMB.pep:\*  
4: /cgn2\_6/prodata/1/pubpaa/US10A\_PUBCOMB.pep:\*  
5: /cgn2\_6/prodata/1/pubpaa/US10B\_PUBCOMB.pep:\*  
6: /cgn2\_6/prodata/1/pubpaa/US11\_PUBCOMB.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	323	100.0	323	3	US-09-816-095-2
2	323	100.0	323	4	US-10-634-905-2
3	197	61.0	197	3	US-09-816-095-5
4	197	61.0	197	4	US-10-634-905-5
5	81	25.1	145	4	US-10-080-254-93
6	81	25.1	145	4	US-10-242-355-637
7	47	14.6	324	3	US-09-816-095-4
8	47	14.6	324	4	US-10-634-905-4
9	13	4.0	40	4	US-10-029-386-33145
10	12	3.7	167	5	US-10-450-763-43912
11	11	3.4	335	4	US-10-698-190-16
12	11	3.4	335	5	US-10-723-860-4238
13	11	3.4	691	5	US-10-450-763-30691
14	10	3.1	201	4	US-10-403-571-34
15	8	2.5	57	4	US-10-425-115-346410
16	8	2.5	80	4	US-10-437-963-139231
17	8	2.5	87	4	US-10-425-115-313878
18	8	2.5	263	5	US-10-849-948-4
19	8	2.5	277	4	US-10-437-963-143548
20	8	2.5	276	4	US-10-437-963-119079
21	8	2.5	392	4	US-10-156-761-9175
22	8	2.5	402	5	US-10-732-923-7823
23	8	2.5	479	6	US-11-097-143-13308
24	8	2.5	612	4	US-10-437-963-139689
25	8	2.5	807	4	US-10-138-221-7
26	8	2.5	840	4	US-10-425-115-204488
27	8	2.5	912	5	US-10-631-467-1543

28	7	2.2	19	3	US-09-864-761-39306	Sequence 39306, A
29	7	2.2	36	3	US-09-864-761-41351	Sequence 41351, A
30	7	2.2	44	4	US-10-242-355-404	Sequence 404, App
31	7	2.2	57	3	US-09-729-674-38	Sequence 38, Appl
32	7	2.2	57	5	US-10-913-553-38	Sequence 38, Appl
33	7	2.2	58	4	US-10-425-115-295950	Sequence 295950, A
34	7	2.2	64	4	US-10-767-701-50698	Sequence 50698, A
35	7	2.2	70	4	US-10-425-115-275941	Sequence 275941, A
36	7	2.2	70	4	US-10-425-115-336480	Sequence 336480, A
37	7	2.2	72	4	US-10-424-599-183280	Sequence 183280, A
38	7	2.2	72	4	US-10-767-701-50207	Sequence 50207, A
39	7	2.2	76	4	US-10-424-599-183282	Sequence 183282, A
40	7	2.2	80	5	US-10-287-436A-603	Sequence 603, App
41	7	2.2	82	5	US-10-450-763-57491	Sequence 57491, A
42	7	2.2	86	4	US-10-425-115-200651	Sequence 200651, A
43	7	2.2	87	3	US-09-882-227-422	Sequence 422, App
44	7	2.2	87	4	US-10-335-977-7681	Sequence 7681, Ap
45	7	2.2	89	4	US-10-437-963-112770	Sequence 112770, A

ALIGNMENTS

RESULT 1  
US-09-816-095-2  
; Sequence 2, Application US/09816095  
; Patent No. US20020137164A1  
; GENERAL INFORMATION:  
; APPLICANT: GAN, Weiniu  
; TITLE OF INVENTION: ISOLATED HUMAN ENZYME PROTEINS, NUCLEIC  
; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN ENZYME PROTEINS, AND USES  
; TITLE OF INVENTION: THEREOF  
; FILE REFERENCE: CL001147  
; CURRENT APPLICATION NUMBER: US/09/816,095  
; CURRENT FILING DATE: 2001-03-26  
; NUMBER OF SEQ ID NOS: 5  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 2  
; LENGTH: 323  
; TYPE: PRT  
; ORGANISM: Human  
US-09-816-095-2

Query Match	100.00%;	Score 323;	DB 3;	Length 323;
Best Local Similarity	100.00%;	Pred. No. 3.2e-300;		
Matches 323;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	1	MKSALTRFFILLPWILVIIMLDVDRTRPVPLTPRPYFSPYAVGGRGARLPLRRGGPA	60	
Db	1	MKSALTRFFILLPWILVIIMLDVDRTRPVPLTPRPYFSPYAVGGRGARLPLRRGGPA	60	
QY	61	HGTQKRNQSRQPOPEPQLPTIYATPTYSRPVQKAELTRLANTFRQVQLHILVEDAA	120	
Db	61	HGTQKRNQSRQPOPEPQLPTIYATPTYSRPVQKAELTRLANTFRQVQLHILVEDAA	120	
QY	121	ARSELVSRFLARAGLPSTHLVPTPRYKRPGLPRATEORNAGLWLRORHQRAPGV	180	
Db	121	ARSELVSRFLARAGLPSTHLVPTPRYKRPGLPRATEORNAGLWLRORHQRAPGV	180	
QY	181	LFADFDDNTYSLELFOEMTTTRKVSVPVGLVGGRRYERPLVGVGKVVGYTTCWRADRP	240	
Db	181	LFADFDDNTYSLELFOEMTTTRKVSVPVGLVGGRRYERPLVGVGKVVGYTTCWRADRP	240	
QY	241	AIDMAGFVSLQVILSNPKAVFKRGSGQGMQESDFLKQITTVVEELPKANNCTKVLVWH	300	
Db	241	AIDMAGFVSLQVILSNPKAVFKRGSGQGMQESDFLKQITTVVEELPKANNCTKVLVWH	300	
QY	301	TRTEKVNLANEPKYHLDTVKIEV	323	
Db	301	TRTEKVNLANEPKYHLDTVKIEV	323	

RESULT 2

```
US-10-634-905-2
; Sequence 2, Application US/10634905
; Publication No. US20040067225A1
; GENERAL INFORMATION:
; APPLICANT: GAN, Weiniu et al.
; TITLE OF INVENTION: ISOLATED HUMAN ENZYME PROTEINS, NUCLEIC
; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN ENZYME PROTEINS, AND USES
; TITLE OF INVENTION: THEREOF
; FILE REFERENCE: CL001147-DIV
; CURRENT FILING DATE: 2003-08-06
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 323
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-634-905-2

Query Match      100.0%; Score 323; DB 4; Length 323;
Best Local Similarity 100.0%; Pred. No. 3.2e-300;
Matches 323; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 M K S A L T R F F I L L P W I L I V I M L D V D T R R P V P L T P R P Y F S P V A V G R G G A R L P L R G G P A 60
D b 1 M K S A L T R F F I L L P W I L I V I M L D V D T R R P V P L T P R P Y F S P V A V G R G G A R L P L R G G P A 60

QY 61 H G T Q K R N Q S R P Q P E P Q L P T T I A I T P T Y S R P V Q K A E L T R L A N T F R Q V A Q L H W I L V E D A A 120
D b 61 H G T Q K R N Q S R P Q P E P Q L P T T I A I T P T Y S R P V Q K A E L T R L A N T F R Q V A Q L H W I L V E D A A 120

QY 121 A R S E L V S R F L A R A G L P S T H L H V P T P R R Y K R P G L P R A T E O R N A G L A W L R O R H O H Q R A O P G V 180
D b 121 A R S E L V S R F L A R A G L P S T H L H V P T P R R Y K R P G L P R A T E O R N A G L A W L R O R H O H Q R A O P G V 180

QY 181 L F F A D D D N T Y S L E L F Q E 197
D b 181 L F F A D D D N T Y S L E L F Q E 197

RESULT 4
US-10-634-905-5
; Sequence 5, Application US/10634905
; Publication No. US20040067225A1
; GENERAL INFORMATION:
; APPLICANT: GAN, Weiniu et al.
; TITLE OF INVENTION: ISOLATED HUMAN ENZYME PROTEINS, NUCLEIC
; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN ENZYME PROTEINS, AND USES
; TITLE OF INVENTION: THEREOF
; FILE REFERENCE: CL001147-DIV
; CURRENT FILING DATE: 2003-08-06
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5
; LENGTH: 197
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-634-905-5

Query Match      61.0%; Score 197; DB 4; Length 197;
Best Local Similarity 100.0%; Pred. No. 7.3e-180;
Matches 197; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 M K S A L T R F F I L L P W I L I V I M L D V D T R R P V P L T P R P Y F S P V A V G R G G A R L P L R G G P A 60
D b 1 M K S A L T R F F I L L P W I L I V I M L D V D T R R P V P L T P R P Y F S P V A V G R G G A R L P L R G G P A 60

QY 61 H G T Q K R N Q S R P Q P E P Q L P T T I A I T P T Y S R P V Q K A E L T R L A N T F R Q V A Q L H W I L V E D A A 120
D b 61 H G T Q K R N Q S R P Q P E P Q L P T T I A I T P T Y S R P V Q K A E L T R L A N T F R Q V A Q L H W I L V E D A A 120

QY 121 A R S E L V S R F L A R A G L P S T H L H V P T P R R Y K R P G L P R A T E O R N A G L A W L R O R H O H Q R A O P G V 180
D b 121 A R S E L V S R F L A R A G L P S T H L H V P T P R R Y K R P G L P R A T E O R N A G L A W L R O R H O H Q R A O P G V 180

QY 181 L F F A D D D N T Y S L E L F Q E 197
D b 181 L F F A D D D N T Y S L E L F Q E 197

RESULT 5
US-10-080-254-93
; Sequence 93, Application US/10080254
; Publication No. US20030199008A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: P114C1
; CURRENT FILING DATE: 2002-02-22
; NUMBER OF SEQ ID NOS: 169
; Prior Application removed - See File Wrapper or Palm
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 93
; LENGTH: 145
; TYPE: PRT

QY 1 M K S A L T R F F I L L P W I L I V I M L D V D T R R P V P L T P R P Y F S P V A V G R G G A R L P L R G G P A 60
D b 1 M K S A L T R F F I L L P W I L I V I M L D V D T R R P V P L T P R P Y F S P V A V G R G G A R L P L R G G P A 60

QY 61 H G T Q K R N Q S R P Q P E P Q L P T T I A I T P T Y S R P V Q K A E L T R L A N T F R Q V A Q L H W I L V E D A A 120
D b 61 H G T Q K R N Q S R P Q P E P Q L P T T I A I T P T Y S R P V Q K A E L T R L A N T F R Q V A Q L H W I L V E D A A 120

QY 121 A R S E L V S R F L A R A G L P S T H L H V P T P R R Y K R P G L P R A T E O R N A G L A W L R O R H O H Q R A O P G V 180
D b 121 A R S E L V S R F L A R A G L P S T H L H V P T P R R Y K R P G L P R A T E O R N A G L A W L R O R H O H Q R A O P G V 180

QY 181 L F F A D D D N T Y S L E L F Q E 197
D b 181 L F F A D D D N T Y S L E L F Q E 197

RESULT 3
US-09-816-095-5
; Sequence 5, Application US/09816095
; Patent No. US20020137164A1
; GENERAL INFORMATION:
; APPLICANT: GAN, Weiniu
; TITLE OF INVENTION: ISOLATED HUMAN ENZYME PROTEINS, NUCLEIC
; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN ENZYME PROTEINS, AND USES
; TITLE OF INVENTION: THEREOF
; FILE REFERENCE: CL001147
; CURRENT FILING DATE: 2001-03-26
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5
; LENGTH: 197
; TYPE: PRT
; ORGANISM: Human
US-09-816-095-5

Query Match      61.0%; Score 197; DB 3; Length 197;
Best Local Similarity 100.0%; Pred. No. 7.3e-180;
Matches 197; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 M K S A L T R F F I L L P W I L I V I M L D V D T R R P V P L T P R P Y F S P V A V G R G G A R L P L R G G P A 60
D b 1 M K S A L T R F F I L L P W I L I V I M L D V D T R R P V P L T P R P Y F S P V A V G R G G A R L P L R G G P A 60

QY 61 H G T Q K R N Q S R P Q P E P Q L P T T I A I T P T Y S R P V Q K A E L T R L A N T F R Q V A Q L H W I L V E D A A 120
D b 61 H G T Q K R N Q S R P Q P E P Q L P T T I A I T P T Y S R P V Q K A E L T R L A N T F R Q V A Q L H W I L V E D A A 120

QY 121 A R S E L V S R F L A R A G L P S T H L H V P T P R R Y K R P G L P R A T E O R N A G L A W L R O R H O H Q R A O P G V 180
D b 121 A R S E L V S R F L A R A G L P S T H L H V P T P R R Y K R P G L P R A T E O R N A G L A W L R O R H O H Q R A O P G V 180

QY 181 L F F A D D D N T Y S L E L F Q E 197
D b 181 L F F A D D D N T Y S L E L F Q E 197

QY 241 A I D M A G F A V S L Q V I L S N P K A V F K R G S Q P G M Q E S D F L K Q I T T V E E L E P K A N N C T K V L V W H 300
D b 241 A I D M A G F A V S L Q V I L S N P K A V F K R G S Q P G M Q E S D F L K Q I T T V E E L E P K A N N C T K V L V W H 300

QY 301 T R T E K V N L A N E P K Y H L D T V K I E V 323
D b 301 T R T E K V N L A N E P K Y H L D T V K I E V 323

RESULT 2
US-09-816-095-5
; Sequence 5, Application US/09816095
; Patent No. US20020137164A1
; GENERAL INFORMATION:
; APPLICANT: GAN, Weiniu
; TITLE OF INVENTION: ISOLATED HUMAN ENZYME PROTEINS, NUCLEIC
; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN ENZYME PROTEINS, AND USES
; TITLE OF INVENTION: THEREOF
; FILE REFERENCE: CL001147
; CURRENT FILING DATE: 2001-03-26
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5
; LENGTH: 197
; TYPE: PRT
; ORGANISM: Human
US-09-816-095-5

Query Match      61.0%; Score 197; DB 3; Length 197;
Best Local Similarity 100.0%; Pred. No. 7.3e-180;
Matches 197; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 M K S A L T R F F I L L P W I L I V I M L D V D T R R P V P L T P R P Y F S P V A V G R G G A R L P L R G G P A 60
D b 1 M K S A L T R F F I L L P W I L I V I M L D V D T R R P V P L T P R P Y F S P V A V G R G G A R L P L R G G P A 60

QY 61 H G T Q K R N Q S R P Q P E P Q L P T T I A I T P T Y S R P V Q K A E L T R L A N T F R Q V A Q L H W I L V E D A A 120
D b 61 H G T Q K R N Q S R P Q P E P Q L P T T I A I T P T Y S R P V Q K A E L T R L A N T F R Q V A Q L H W I L V E D A A 120

QY 121 A R S E L V S R F L A R A G L P S T H L H V P T P R R Y K R P G L P R A T E O R N A G L A W L R O R H O H Q R A O P G V 180
D b 121 A R S E L V S R F L A R A G L P S T H L H V P T P R R Y K R P G L P R A T E O R N A G L A W L R O R H O H Q R A O P G V 180

QY 181 L F F A D D D N T Y S L E L F Q E 197
D b 181 L F F A D D D N T Y S L E L F Q E 197

QY 241 A I D M A G F A V S L Q V I L S N P K A V F K R G S Q P G M Q E S D F L K Q I T T V E E L E P K A N N C T K V L V W H 300
D b 241 A I D M A G F A V S L Q V I L S N P K A V F K R G S Q P G M Q E S D F L K Q I T T V E E L E P K A N N C T K V L V W H 300

QY 301 T R T E K V N L A N E P K Y H L D T V K I E V 323
D b 301 T R T E K V N L A N E P K Y H L D T V K I E V 323
```

```
; ORGANISM: Homo sapiens
US-10-080-254-93

Query Match      25.1%; Score 81; DB 4; Length 145;
Best Local Similarity 100.0%; Pred. No. 5.4e-69;
Matches 81; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 210 GLVGGRRYERPLVNGKVVGWYTCWRADRPFAIDMAGFAVSLQVILSNPKAVFKRRGSQP 269
      |||
Db 23 GLVGGRRYERPLVNGKVVGWYTCWRADRPFAIDMAGFAVSLQVILSNPKAVFKRRGSQP 82

Qy 270 GQESDFLKQITTVEELEPKA 290
      |||
Db 83 GQESDFLKQITTVEELEPKA 103

RESULT 6
US-10-242-355-637
; Sequence 637, Application US/10242355
; Publication No. US20030235831A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC003C1
; CURRENT APPLICATION NUMBER: US/10/242,355
; CURRENT FILING DATE: 2002-09-13
; PRIOR APPLICATION NUMBER: 09/764,897
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: 60/179,065
; PRIOR FILING DATE: 2000-01-31
; PRIOR APPLICATION NUMBER: 60/180,628
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: 60/214,886
; PRIOR FILING DATE: 2000-06-28
; PRIOR APPLICATION NUMBER: 60/217,487
; PRIOR FILING DATE: 2000-07-11
; PRIOR APPLICATION NUMBER: 60/225,758
; PRIOR FILING DATE: 2000-08-14
; PRIOR APPLICATION NUMBER: 60/220,963
; PRIOR FILING DATE: 2000-07-26
; PRIOR APPLICATION NUMBER: 60/217,496
; PRIOR FILING DATE: 2000-07-11
; PRIOR APPLICATION NUMBER: 60/225,447
; PRIOR FILING DATE: 2000-08-14
; PRIOR APPLICATION NUMBER: 60/218,290
; PRIOR FILING DATE: 2000-07-14
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 1267
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 637
; LENGTH: 145
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-242-355-637

Query Match      25.1%; Score 81; DB 4; Length 145;
Best Local Similarity 100.0%; Pred. No. 5.4e-69;
Matches 81; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 210 GLVGGRRYERPLVNGKVVGWYTCWRADRPFAIDMAGFAVSLQVILSNPKAVFKRRGSQP 269
      |||
Db 23 GLVGGRRYERPLVNGKVVGWYTCWRADRPFAIDMAGFAVSLQVILSNPKAVFKRRGSQP 82

Qy 270 GQESDFLKQITTVEELEPKA 290
      |||
Db 83 GQESDFLKQITTVEELEPKA 103

RESULT 7
US-09-816-095-4
; Sequence 4, Application US/09816095
; Patent No. US20020137164A1
; GENERAL INFORMATION:
```

```
; APPLICANT: GAN, Weiniu
; TITLE OF INVENTION: ISOLATED HUMAN ENZYME PROTEINS, NUCLEIC
; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN ENZYME PROTEINS, AND USES
; FILE REFERENCE: CL001147
; CURRENT APPLICATION NUMBER: US/09/816,095
; CURRENT FILING DATE: 2001-03-26
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 324
; TYPE: PRT
; ORGANISM: Rattus norvegicus
US-09-816-095-4

Query Match      14.6%; Score 47; DB 3; Length 324;
Best Local Similarity 100.0%; Pred. No. 3.6e-36;
Matches 47; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 176 AQPGLVFFADDDNTYSLELFPQEMRTTRKVSVPVGLVGGRRYERPLV 222
      |||
Db 177 AQPGLVFFADDDNTYSLELFPQEMRTTRKVSVPVGLVGGRRYERPLV 223

RESULT 8
US-10-634-905-4
; Sequence 4, Application US/10634905
; Publication No. US20040067225A1
; GENERAL INFORMATION:
; APPLICANT: GAN, Weiniu et al.
; TITLE OF INVENTION: ISOLATED HUMAN ENZYME PROTEINS, NUCLEIC
; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN ENZYME PROTEINS, AND USES
; FILE REFERENCE: CL001147-DIV
; CURRENT APPLICATION NUMBER: US/10/634,905
; CURRENT FILING DATE: 2003-08-06
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 324
; TYPE: PRT
; ORGANISM: Rattus norvegicus
US-10-634-905-4

Query Match      14.6%; Score 47; DB 4; Length 324;
Best Local Similarity 100.0%; Pred. No. 3.6e-36;
Matches 47; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 176 AQPGLVFFADDDNTYSLELFPQEMRTTRKVSVPVGLVGGRRYERPLV 222
      |||
Db 177 AQPGLVFFADDDNTYSLELFPQEMRTTRKVSVPVGLVGGRRYERPLV 223

RESULT 9
US-10-029-386-33145
; Sequence 33145, Application US/10029386
; Publication No. US20030194704A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR G
; TITLE OF INVENTION: EXPRESSION ANALYSIS TWO
; FILE REFERENCE: AEOMICA-X-2
; CURRENT APPLICATION NUMBER: US/10/029,386
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 34288
; SOFTWARE: Annonmax Sequence Listing Engine vers. 1.1
; SEQ ID NO 33145
; LENGTH: 40
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
```

; OTHER INFORMATION: MAP TO AC013591.3  
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.5  
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.6  
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.6  
; OTHER INFORMATION: SWISSPROT HIT: Q9P2W7, EVALUAE 9.00e-18  
US-10-029-386-33145

Query Match 4.0%; Score 13; DB 4; Length 40;  
Best Local Similarity 100.0%; Pred. No. 0.0002;  
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 183 FADDNTYSLELF 195  
|||||

DB 26 FADDNTYSLELF 38  
|||||

RESULT 10  
US-10-450-763-43912  
; Sequence 43912, Application US/10450763  
; Publication No. US20050196754A1  
; GENERAL INFORMATION:  
; APPLICANT: Hyseq, Inc  
; TITLE OF INVENTION: NOVEL NUCLEIC ACIDS AND POLYPEPTIDES  
; FILE REFERENCE: 790CIP3/US  
; CURRENT APPLICATION NUMBER: US/10/450,763  
; CURRENT FILING DATE: 2003-06-11  
; PRIOR APPLICATION NUMBER: PCT/US01/08631  
; PRIOR FILING DATE: 2001-03-30  
; PRIOR APPLICATION NUMBER: 09/540,217  
; PRIOR FILING DATE: 2000-03-31  
; PRIOR APPLICATION NUMBER: 09/649,167  
; PRIOR FILING DATE: 2000-08-23  
; NUMBER OF SEQ ID NOS: 60736  
; SOFTWARE: Custom  
; SEQ ID NO 43912  
; LENGTH: 167  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-450-763-43912

Query Match 3.7%; Score 12; DB 5; Length 167;  
Best Local Similarity 100.0%; Pred. No. 0.0062;  
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 238 RPFAIDMAGFAV 249  
|||||

DB 42 RPFAIDMAGFAV 53  
|||||

RESULT 11  
US-10-698-190-16  
; Sequence 16, Application US/10698190  
; Publication No. US20040147469A1  
; GENERAL INFORMATION:  
; APPLICANT: Silver, J.  
; APPLICANT: Grimpe, B.  
; TITLE OF INVENTION: Methods of Inhibiting Glial Scar Formation  
; FILE REFERENCE: CWRU-P01-018  
; CURRENT APPLICATION NUMBER: US/10/698,190  
; CURRENT FILING DATE: 2003-10-31  
; NUMBER OF SEQ ID NOS: 40  
; SOFTWARE: Patentin version 3.1  
; SEQ ID NO 16  
; LENGTH: 335  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-698-190-16

Query Match 3.4%; Score 11; DB 4; Length 335;  
Best Local Similarity 100.0%; Pred. No. 0.1;  
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 204 VSWVPVGLVGG 214  
|||||

DB 213 VSWVPVGLVGG 223  
|||||

RESULT 12  
US-10-723-860-4238  
; Sequence 4238, Application US/10723860  
; Publication No. US20040253606A1  
; GENERAL INFORMATION:  
; APPLICANT: Aziz, Natasha  
; APPLICANT: Ginsburg, Wendy M.  
; APPLICANT: Zlotnik, Albert  
; TITLE OF INVENTION: Methods of Diagnosis of Soft Tissue Sarcoma, Compositions &  
; TITLE OF INVENTION: Methods for Screening for Soft Tissue Sarcoma Modulators  
; FILE REFERENCE: 05882.0193.NPUS01  
; CURRENT APPLICATION NUMBER: US/10/723,860  
; CURRENT FILING DATE: 2003-11-26  
; PRIOR APPLICATION NUMBER: 60/429,739  
; PRIOR FILING DATE: 2002-11-26  
; NUMBER OF SEQ ID NOS: 8393  
; SOFTWARE: Patentin version 3.2  
; SEQ ID NO 4238  
; LENGTH: 335  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-723-860-4238

Query Match 3.4%; Score 11; DB 5; Length 335;  
Best Local Similarity 100.0%; Pred. No. 0.1;  
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 204 VSWVPVGLVGG 214  
|||||

DB 213 VSWVPVGLVGG 223  
|||||

RESULT 13  
US-10-450-763-30691  
; Sequence 30691, Application US/10450763  
; Publication No. US20050196754A1  
; GENERAL INFORMATION:  
; APPLICANT: Hyseq, Inc  
; TITLE OF INVENTION: NOVEL NUCLEIC ACIDS AND POLYPEPTIDES  
; FILE REFERENCE: 790CIP3/US  
; CURRENT APPLICATION NUMBER: US/10/450,763  
; CURRENT FILING DATE: 2003-06-11  
; PRIOR APPLICATION NUMBER: PCT/US01/08631  
; PRIOR FILING DATE: 2001-03-30  
; PRIOR APPLICATION NUMBER: 09/540,217  
; PRIOR FILING DATE: 2000-03-31  
; PRIOR APPLICATION NUMBER: 09/649,167  
; PRIOR FILING DATE: 2000-08-23  
; NUMBER OF SEQ ID NOS: 60736  
; SOFTWARE: Custom  
; SEQ ID NO 30691  
; LENGTH: 691  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: DOMAIN  
; LOCATION: (583)..(632)  
; OTHER INFORMATION: PHOSPHORYLASE KINASE ALPHA MUSCL domain identified by  
; OTHER INFORMATION: eMATRIX, accession number PD01841F, p-value=7.048e-09, raw score  
; OTHER INFORMATION: 13.36  
US-10-450-763-30691

Query Match 3.4%; Score 11; DB 5; Length 691;  
Best Local Similarity 100.0%; Pred. No. 0.19;  
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 204 VSWVPVGLVGG 214  
|||||

DB 348 VSWVPVGLVGG 358  
|||||

```
RESULT 14
US-10-403-571-34
; Sequence 34, Application US/10403571
; Publication No. US20040068763A1
; GENERAL INFORMATION:
; APPLICANT: Hopkins, Nancy
; APPLICANT: Golling, Gregory
; APPLICANT: Amsterdam, Adam
; APPLICANT: Sun, Zhaoxia
; TITLE OF INVENTION: Developmental Mutations in Zebrafish
; FILE REFERENCE: 01997/539002
; CURRENT APPLICATION NUMBER: US/10/403,571
; CURRENT FILING DATE: 2003-03-25
; PRIOR APPLICATION NUMBER: US 60/368,760
; PRIOR FILING DATE: 2002-03-29
; NUMBER OF SEQ ID NOS: 159
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 34
; LENGTH: 201
; TYPE: PRT
; ORGANISM: Danio rerio
US-10-403-571-34
Query Match      3.1%; Score 10; DB 4; Length 201;
Best Local Similarity 100.0%; Pred. No. 0.6;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 183 FADDNTYSL 192
Db 187 FADDNTYSL 196
|||||

RESULT 15
US-10-425-115-346410
; Sequence 346410, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 346410
; LENGTH: 57
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_79090C.1.pep
US-10-425-115-346410
Query Match      2.5%; Score 8; DB 4; Length 57;
Best Local Similarity 100.0%; Pred. No. 17;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 133 AGLPSTHL 140
Db 47 AGLPSTHL 54
|||||

Search completed: May 9, 2006, 11:00:09
Job time : 79 secs
```

**This Page Blank (uspto)**

GenCore version 5.1.8  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: May 9, 2006, 10:59:55 ; Search time 12 Seconds  
(without alignments)  
1245.831 Million cell updates/sec

Title: US-10-634-905-2  
Perfect score: 323  
Sequence: 1 MKSALFTRFFILLPWILVI.....EKVLANEPKYHLDTVKIEV 323

Scoring table: OLIGO  
Gapop 60.0 , Gapext 60.0

Searched: 235405 seqs, 46284737 residues

Word size : 1

Total number of hits satisfying chosen parameters: 235171

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Published Applications\_AA\_New:\*  
1: /SID55/ptodata/1/pubpaa/US08\_NEW\_PUB.pep.\*  
2: /SID55/ptodata/1/pubpaa/US06\_NEW\_PUB.pep.\*  
3: /SID55/ptodata/1/pubpaa/US07\_NEW\_PUB.pep.\*  
4: /SID55/ptodata/1/pubpaa/US08\_NEW\_PUB.pep.\*  
5: /SID55/ptodata/1/pubpaa/PCT\_NEW\_PUB.pep.\*  
6: /SID55/ptodata/1/pubpaa/US09\_NEW\_PUB.pep.\*  
7: /SID55/ptodata/1/pubpaa/US09\_NEW\_PUB.pep.\*  
8: /SID55/ptodata/1/pubpaa/US10\_NEW\_PUB.pep.\*  
9: /SID55/ptodata/1/pubpaa/US10\_NEW\_PUB.pep.\*  
10: /SID55/ptodata/1/pubpaa/US11\_NEW\_PUB.pep.\*  
11: /SID55/ptodata/1/pubpaa/US11\_NEW\_PUB.pep.\*  
12: /SID55/ptodata/1/pubpaa/US60\_NEW\_PUB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	323	100.0	323	9	US-10-330-773-566
2	53	16.4	290	9	US-10-330-773-563
3	8	2.5	498	9	US-10-467-657-5006
4	7	2.2	107	11	US-11-264-096-871
5	7	2.2	144	11	US-11-096-568A-26178
6	7	2.2	160	11	US-11-096-568A-26177
7	7	2.2	199	11	US-11-079-463-9881
8	7	2.2	205	11	US-11-113-424-42
9	7	2.2	212	11	US-11-188-298-12238
10	7	2.2	237	11	US-11-188-298-3599
11	7	2.2	237	11	US-11-188-298-4760
12	7	2.2	237	11	US-11-188-298-12773
13	7	2.2	258	11	US-11-096-568A-182
14	7	2.2	263	11	US-11-072-512-2192
15	7	2.2	293	11	US-11-096-568A-16425
16	7	2.2	295	11	US-11-096-568A-30519
17	7	2.2	319	11	US-11-087-099-2924
18	7	2.2	322	11	US-11-096-568A-30518
19	7	2.2	331	11	US-11-188-298-18650
20	7	2.2	358	9	US-10-793-626-18
21	7	2.2	358	11	US-11-188-298-14936

22 7 2.2 387 11 US-11-098-686-11142 Sequence 11142, A  
23 7 2.2 408 11 US-11-140-417-6 Sequence 6, Appl  
24 7 2.2 428 11 US-11-229-484-16 Sequence 16, Appl  
25 7 2.2 472 11 US-11-087-099-10910 Sequence 10910, A  
26 7 2.2 472 11 US-11-188-298-21080 Sequence 21080, A  
27 7 2.2 488 11 US-11-079-463-6537 Sequence 6537, Ap  
28 7 2.2 569 11 US-11-096-568A-30517 Sequence 30517, A  
29 7 2.2 600 11 US-11-079-463-9991 Sequence 9991, Ap  
30 7 2.2 842 11 US-11-188-298-6227 Sequence 6227, Ap  
31 7 2.2 1574 9 US-10-055-877-211 Sequence 211, App  
32 6 1.9 17 11 US-11-202-449-3 Sequence 3, Appli  
33 6 1.9 33 9 US-10-467-657-8872 Sequence 8872, Ap  
34 6 1.9 34 11 US-11-207-078-318 Sequence 318, App  
35 6 1.9 39 11 US-11-088-785-24 Sequence 24, Appl  
36 6 1.9 42 11 US-11-088-785-23 Sequence 23, Appl  
37 6 1.9 52 11 US-11-088-785-21 Sequence 21, Appl  
38 6 1.9 54 11 US-11-088-785-22 Sequence 22, Appl  
39 6 1.9 71 9 US-10-467-657-6652 Sequence 6652, Ap  
40 6 1.9 72 9 US-10-508-307-7 Sequence 7, Appl  
41 6 1.9 76 9 US-10-821-234-910 Sequence 910, App  
42 6 1.9 83 9 US-10-510-386-114 Sequence 114, App  
43 6 1.9 83 9 US-10-895-064-2389 Sequence 2389, Ap  
44 6 1.9 83 11 US-11-129-741-2389 Sequence 2389, Ap  
45 6 1.9 96 9 US-10-475-075-809 Sequence 809, App

#### ALIGNMENTS

RESULT 1  
US-10-330-773-566  
; Sequence 566, Application US/10330773  
; Publication No. US20060040262A1  
; GENERAL INFORMATION:  
; APPLICANT: David W. Morris  
; APPLICANT: Marc Malandro  
; TITLE OF INVENTION: Novel Compositions and Methods in Cancer  
; FILE REFERENCE: 529452001300  
; CURRENT APPLICATION NUMBER: US/10/330,773  
; CURRENT FILING DATE: 2002-12-27  
; NUMBER OF SEQ ID NOS: 981  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 566  
; LENGTH: 323  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-330-773-566

Query Match 100.0%; Score 323; DB 9; Length 323;  
Best Local Similarity 100.0%; Pred. No. 2.2e-317;  
Matches 323; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MKSALFTRFFILLPWILVIIMLDVDTTRRPVPLTPRPYPSPYAVGEGGARLRRGGPA 60  
Db 1 MKSALFTRFFILLPWILVIIMLDVDTTRRPVPLTPRPYPSPYAVGEGGARLRRGGPA 60  
Qy 61 HGTQKNQSRPQPPQLPTIYAITPTYSRPVQKAELTRANTERQVAOLHILVEDAA 120  
Db 61 HGTQKNQSRPQPPQLPTIYAITPTYSRPVQKAELTRANTERQVAOLHILVEDAA 120  
Qy 121 ARSELVSRFLARAGLPSLTHLVPTPRRYKRPGLPRATEQRNAGLWLRQRHQRAPGV 180  
Db 121 ARSELVSRFLARAGLPSLTHLVPTPRRYKRPGLPRATEQRNAGLWLRQRHQRAPGV 180  
Qy 181 LFPADDNTYSLFQEMRTTRKVSVPVGLVGGRRYERPLVNGKVGVYTGWRADRPF 240  
Db 181 LFPADDNTYSLFQEMRTTRKVSVPVGLVGGRRYERPLVNGKVGVYTGWRADRPF 240  
Qy 241 AIDMAGFVSLQVILSNPKAVFKRGSCQMQSDFLKQITTVVEELEPKANNCTKVLVWH 300  
Db 241 AIDMAGFVSLQVILSNPKAVFKRGSCQMQSDFLKQITTVVEELEPKANNCTKVLVWH 300  
Qy 301 TRTEKVNLANEPKYHLDTVKIEV 323

```
Db      301 TRTEKVNLANEPKXHLDTVKIEV 323
|||||
RESULT 2
US-10-330-773-563
; Sequence 563, Application US/10330773
; Publication No. US20060040262A1
; GENERAL INFORMATION:
; APPLICANT: David W. Morris
; APPLICANT: Marc Malandro
; TITLE OF INVENTION: Novel Compositions and Methods in Cancer
; FILE REFERENCE: 529452001300
; CURRENT APPLICATION NUMBER: US/10/330,773
; CURRENT FILING DATE: 2002-12-27
; NUMBER OF SEQ ID NOS: 981
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 563
; LENGTH: 290
; TYPE: PRT
; ORGANISM: Mus musculus
US-10-330-773-563

Query Match      16.4%; Score 53; DB 9; Length 290;
Best Local Similarity 100.0%; Pred. No. 2.2e-45;
Matches 53; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      237 DRPFAIDMAGFAVSLQVILSNKAVFKRRGSGPQMGSDFLKQITTVESLEPK 289
|||||
Db      206 DRPFAIDMAGFAVSLQVILSNKAVFKRRGSGPQMGSDFLKQITTVESLEPK 258
|||||

RESULT 3
US-10-467-657-5006
; Sequence 5006, Application US/10467657
; Publication No. US20050260581A1
; GENERAL INFORMATION:
; APPLICANT: CHIRON SPA
; APPLICANT: FONTANA Maria Rita
; APPLICANT: PIZZA Mariagrazia
; APPLICANT: MASIGNANI Vega
; APPLICANT: MONACI Elisabetta
; TITLE OF INVENTION: GONOCOCCAL PROTEINS AND NUCLEIC ACIDS
; FILE REFERENCE:
; CURRENT APPLICATION NUMBER: US/10/467,657
; CURRENT FILING DATE: 2003-08-11
; PRIOR APPLICATION NUMBER: GB-0103424.8
; PRIOR FILING DATE: 2001-02-12
; NUMBER OF SEQ ID NOS: 9218
; SOFTWARE: SeqWin99, version 1.04
; SEQ ID NO 5006
; LENGTH: 498
; TYPE: PRT
; ORGANISM: Neisseria gonorrhoeae
US-10-467-657-5006

Query Match      2.5%; Score 8; DB 9; Length 498;
Best Local Similarity 100.0%; Pred. No. 8;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      222 VENGKVVG 229
|||||
Db      136 VENGKVVG 143
|||||

RESULT 4
US-11-264-096-871
; Sequence 871, Application US/11264096
; Publication No. US20060084794A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Albumin Fusion Proteins
; FILE REFERENCE: PF546D1
```

```
; CURRENT APPLICATION NUMBER: US/11/264,096
; CURRENT FILING DATE: 2005-11-02
; PRIOR APPLICATION NUMBER: 09/833,245
; PRIOR FILING DATE: 2001-04-12
; PRIOR APPLICATION NUMBER: 60/229,358
; PRIOR FILING DATE: 2000-04-12
; PRIOR APPLICATION NUMBER: 60/256,931
; PRIOR FILING DATE: 2000-12-21
; PRIOR APPLICATION NUMBER: 60/199,384
; PRIOR FILING DATE: 2000-04-25
; NUMBER OF SEQ ID NOS: 2267
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 871
; LENGTH: 107
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (66)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-11-264-096-871

Query Match      2.2%; Score 7; DB 11; Length 107;
Best Local Similarity 100.0%; Pred. No. 19;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      46 GRGGARL 52
|||||
Db      101 GRGGARL 107
|||||

RESULT 5
US-11-096-568A-26178
; Sequence 26178, Application US/11096568A
; Publication No. US20060048240A1
; GENERAL INFORMATION:
; APPLICANT: Alexandrov, Nikolai et al.
; TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypeptides
; FILE REFERENCE: 2750-1592PUS2
; CURRENT APPLICATION NUMBER: US/11/096,568A
; CURRENT FILING DATE: 2005-04-01
; NUMBER OF SEQ ID NOS: 34471
; SEQ ID NO 26178
; LENGTH: 144
; TYPE: PRT
; ORGANISM: Zea mays subsp. mays
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(144)
; OTHER INFORMATION: Ceres Seq. ID no. 13498759
US-11-096-568A-26178

Query Match      2.2%; Score 7; DB 11; Length 144;
Best Local Similarity 100.0%; Pred. No. 25;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      53 PLRRGGP 59
|||||
Db      59 PLRRGGP 65
|||||

RESULT 6
US-11-096-568A-26177
; Sequence 26177, Application US/11096568A
; Publication No. US20060048240A1
; GENERAL INFORMATION:
; APPLICANT: Alexandrov, Nikolai et al.
; TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypeptides
; FILE REFERENCE: 2750-1592PUS2
; CURRENT APPLICATION NUMBER: US/11/096,568A
; CURRENT FILING DATE: 2005-04-01
```



```
; NUMBER OF SEQ ID NOS: 34471
; SEQ ID NO 26177
; LENGTH: 160
; TYPE: PRT
; ORGANISM: Zea mays subsp. mays
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(160)
; OTHER INFORMATION: Ceres Seq. ID no. 13498758
US-11-096-568A-26177

Query Match          2.2%; Score 7; DB 11; Length 160;
Best Local Similarity 100.0%; Pred. No. 27;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 53 PLRRGGP 59
Db 75 PLRRGGP 81

RESULT 7
US-11-079-463-9881
; Sequence 9881, Application US/11079463
; Publication No. US20060073161A1
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO BACTEROIDES FRATRIUM
; FILE REFERENCE: PATH00-03DIV2
; CURRENT APPLICATION NUMBER: US/11/079,463
; CURRENT FILING DATE: 2005-03-14
; PRIOR APPLICATION NUMBER: US 60/128,705
; PRIOR FILING DATE: 1999-04-09
; PRIOR APPLICATION NUMBER: US 09/540,209
; PRIOR FILING DATE: 2000-04-04
; NUMBER OF SEQ ID NOS: 10444
; SEQ ID NO 9881
; LENGTH: 199
; TYPE: PRT
; ORGANISM: B. fragilis
US-11-079-463-9881

Query Match          2.2%; Score 7; DB 11; Length 199;
Best Local Similarity 100.0%; Pred. No. 34;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 96 AELTRLA 102
Db 4 AELTRLA 10

RESULT 8
US-11-113-424-42
; Sequence 42, Application US/11113424
; Publication No. US20050260713A1
; GENERAL INFORMATION:
; APPLICANT: Gangolli et al.
; TITLE OF INVENTION: Polypeptides and Nucleic Acids Encoding Same
; FILE REFERENCE: 21402-225
; CURRENT APPLICATION NUMBER: US/11/113,424
; CURRENT FILING DATE: 2005-04-21
; PRIOR APPLICATION NUMBER: 60/256,704
; PRIOR FILING DATE: 2000-12-19
; PRIOR APPLICATION NUMBER: 60/311,590
; PRIOR FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: 60/257,314
; PRIOR FILING DATE: 2000-12-20
; PRIOR APPLICATION NUMBER: 60/311,613
; PRIOR FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: 60/315,617
; PRIOR FILING DATE: 2001-08-29
; PRIOR APPLICATION NUMBER: 60/307,506
; PRIOR FILING DATE: 2001-07-24
```

```
; PRIOR APPLICATION NUMBER: 60/322,358
; PRIOR FILING DATE: 2001-09-14
; PRIOR APPLICATION NUMBER: 60/294,075
; PRIOR FILING DATE: 2001-05-29
; PRIOR APPLICATION NUMBER: 60/288,153
; PRIOR FILING DATE: 2001-05-02
; NUMBER OF SEQ ID NOS: 190
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 42
; LENGTH: 205
; TYPE: PRT
; ORGANISM: Mus musculus
US-11-113-424-42

Query Match          2.2%; Score 7; DB 11; Length 205;
Best Local Similarity 100.0%; Pred. No. 35;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 51 RLPLRRG 57
Db 152 RLPLRRG 158

RESULT 9
US-11-188-298-12238
; Sequence 12238, Application US/11188298
; Publication No. US20060075522A1
; GENERAL INFORMATION:
; APPLICANT: Abad, Mark S. et al.
; TITLE OF INVENTION: GENES AND USES FOR PLANT IMPROVEMENT
; FILE REFERENCE: 38-21(53452)B
; CURRENT APPLICATION NUMBER: US/11/188,298
; CURRENT FILING DATE: 2005-07-22
; PRIOR APPLICATION NUMBER: 60/592,978
; PRIOR FILING DATE: 2004-07-31
; NUMBER OF SEQ ID NOS: 22569
; SEQ ID NO 12238
; LENGTH: 212
; TYPE: PRT
; ORGANISM: Paenibacillus macerans
US-11-188-298-12238

Query Match          2.2%; Score 7; DB 11; Length 212;
Best Local Similarity 100.0%; Pred. No. 36;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 58 GPAHGQTQ 64
Db 94 GPAHGQTQ 100

RESULT 10
US-11-188-298-3589
; Sequence 3589, Application US/11188298
; Publication No. US20060075522A1
; GENERAL INFORMATION:
; APPLICANT: Abad, Mark S. et al.
; TITLE OF INVENTION: GENES AND USES FOR PLANT IMPROVEMENT
; FILE REFERENCE: 38-21(53452)B
; CURRENT APPLICATION NUMBER: US/11/188,298
; CURRENT FILING DATE: 2005-07-22
; PRIOR APPLICATION NUMBER: 60/592,978
; PRIOR FILING DATE: 2004-07-31
; NUMBER OF SEQ ID NOS: 22569
; SEQ ID NO 3589
; LENGTH: 237
; TYPE: PRT
; ORGANISM: Paenibacillus macerans
US-11-188-298-3589

Query Match          2.2%; Score 7; DB 11; Length 237;
Best Local Similarity 100.0%; Pred. No. 40;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

QY 58 GPAHGQTQ 64  
|  
Db 119 GPAHGQTQ 125

## RESULT 11

US-11-188-298-4760  
; Sequence 4760, Application US/11188298  
; Publication No. US20060075522A1

## ; GENERAL INFORMATION:

; APPLICANT: Abad, Mark S. et al.  
; TITLE OF INVENTION: GENES AND USES FOR PLANT IMPROVEMENT  
; FILE REFERENCE: 38-21(53452)B  
; CURRENT APPLICATION NUMBER: US/11/188,298  
; CURRENT FILING DATE: 2005-07-22  
; PRIOR APPLICATION NUMBER: 60/592,978  
; PRIOR FILING DATE: 2004-07-31  
; NUMBER OF SEQ ID NOS: 22569  
; SEQ ID NO 4760  
; LENGTH: 237  
; TYPE: PRT  
; ORGANISM: Paenibacillus macerans  
US-11-188-298-4760

Query Match 2.2%; Score 7; DB 11; Length 237;  
Best Local Similarity 100.0%; Pred. No. 40;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 58 GPAHGQTQ 64  
|  
Db 119 GPAHGQTQ 125

## RESULT 12

US-11-188-298-12773  
; Sequence 12773, Application US/11188298  
; Publication No. US20060075522A1

## ; GENERAL INFORMATION:

; APPLICANT: Abad, Mark S. et al.  
; TITLE OF INVENTION: GENES AND USES FOR PLANT IMPROVEMENT  
; FILE REFERENCE: 38-21(53452)B  
; CURRENT APPLICATION NUMBER: US/11/188,298  
; CURRENT FILING DATE: 2005-07-22  
; PRIOR APPLICATION NUMBER: 60/592,978  
; PRIOR FILING DATE: 2004-07-31  
; NUMBER OF SEQ ID NOS: 22569  
; SEQ ID NO 12773  
; LENGTH: 237  
; TYPE: PRT  
; ORGANISM: Paenibacillus macerans  
US-11-188-298-12773

Query Match 2.2%; Score 7; DB 11; Length 237;  
Best Local Similarity 100.0%; Pred. No. 40;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 58 GPAHGQTQ 64  
|  
Db 119 GPAHGQTQ 125

## RESULT 13

US-11-096-568A-182  
; Sequence 182, Application US/11096568A  
; Publication No. US20060048240A1

## ; GENERAL INFORMATION:

; APPLICANT: Alexandrov, Nikolai et al.  
; TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypeptides  
; FILE REFERENCE: 2750-1592PGS2  
; CURRENT APPLICATION NUMBER: US/11/096,568A  
; CURRENT FILING DATE: 2005-04-01

; NUMBER OF SEQ ID NOS: 34471  
; SEQ ID NO 182  
; LENGTH: 258  
; TYPE: PRT  
; ORGANISM: Zea mays subsp. mays  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (1)..(258)  
; OTHER INFORMATION: Ceres Seq. ID no. 14305256  
US-11-096-568A-182

Query Match 2.2%; Score 7; DB 11; Length 258;  
Best Local Similarity 100.0%; Pred. No. 43;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 209 VGLVGGR 215  
|  
Db 140 VGLVGGR 146

## RESULT 14

US-11-072-512-2192  
; Sequence 2192, Application US/11072512  
; Publication No. US20060029945A1

## ; GENERAL INFORMATION:

; APPLICANT: ISOGAI, TAKAO  
; APPLICANT: SUGIYAMA, TOMOYASU  
; APPLICANT: OTSUKI, TETSUJI  
; APPLICANT: WAKAMATSU, AI  
; APPLICANT: SATO, HIROYUKI  
; APPLICANT: ISHII, SHIZUKO  
; APPLICANT: YAMAMOTO, JUN-ICHI  
; APPLICANT: ISONO, YUUKO  
; APPLICANT: HIO, YURI  
; APPLICANT: OTSUKA, KAORU  
; APPLICANT: NAGAI, KEIICHI  
; APPLICANT: IRIE, KYOTARO  
; APPLICANT: TAMECHIKA, ICHIRO  
; APPLICANT: SEKI, NAOHICO  
; APPLICANT: YOSHIKAWA, TSUTOMU  
; APPLICANT: OTSUKA, MOTYUKI  
; APPLICANT: NAGAHARI, KENJI  
; APPLICANT: MASUHO, YASUHIKO  
; TITLE OF INVENTION: Novel full length cDNA  
; FILE REFERENCE: 084335-0191  
; CURRENT APPLICATION NUMBER: US/11/072,512  
; CURRENT FILING DATE: 2005-03-07  
; PRIOR APPLICATION NUMBER: US 60/350,978  
; PRIOR FILING DATE: 2002-01-25  
; PRIOR APPLICATION NUMBER: JP 2001-379298  
; PRIOR FILING DATE: 2001-11-05  
; NUMBER OF SEQ ID NOS: 4096  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 2192  
; LENGTH: 263  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-11-072-512-2192

Query Match 2.2%; Score 7; DB 11; Length 263;  
Best Local Similarity 100.0%; Pred. No. 44;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 211 LVGGRRY 217  
|  
Db 189 LVGGRRY 195

## RESULT 15

US-11-096-568A-16425  
; Sequence 16425, Application US/11096568A  
; Publication No. US20060048240A1  
; GENERAL INFORMATION:

; APPLICANT: Alexandrov, Nickolai et al.  
; TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypeptides  
; FILE OF INVENTION: Therby  
; FILE REFERENCE: 2750-1592PUS2  
; CURRENT APPLICATION NUMBER: US/11/096,568A  
; CURRENT FILING DATE: 2005-04-01  
; NUMBER OF SEQ ID NOS: 34471  
; SEQ ID NO 16425  
; LENGTH: 293  
; TYPE: PRT  
; ORGANISM: Zea mays subsp. mays  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (1)..(293)  
; OTHER INFORMATION: Ceres Seq. ID no. 12352233  
US-11-096-568A-16425

Query Match 2.2%; Score 7; DB 11; Length 293;  
Best Local Similarity 100.0%; Pred. No. 49;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 45 VGRGGAR 51  
| | | | |  
Db 69 VGRGGAR 75

Search completed: May 9, 2006, 11:00:26  
Job time : 12 secs

**This Page Blank (uspto)**

GenCore version 5.1.8  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: May 9, 2006, 10:51:59 ; Search time 110 Seconds  
(without alignments)  
2071.688 Million cell updates/sec

Title: US-10-634-905-2

Perfect score: 1705

Sequence: 1 MKSALFTRFPILLPILVI.....EKVNLANEKPYHLDTVKIEV 323

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2166443 seqs, 705528306 residues

Total number of hits satisfying chosen parameters: 2166443

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

UniProt\_05\_80.\*

1: uniprot\_sprot.\*

2: uniprot\_trembl.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1705	100.0	323	1 B3GA2_HUMAN	Q9npz5 homo sapien
2	1705	100.0	323	2 Q599K0_PANTR	Q599k0 pan troglod
3	1620	95.0	329	1 B3GA2_CANFA	O5caz6 canis famil
4	1592.5	93.4	326	2 Q599K1_BOVIN	Q599k1 bos taurus
5	1531.5	89.8	324	1 B3GA2_MOUSE	F59270 mus musculu
6	1531.5	89.5	324	2 Q5DTT7_MOUSE	Q5dt7 mus musculu
7	1526.5	89.5	324	1 B3GA2_RAT	Q9z137 rattus norv
8	1221	71.6	331	2 Q599J8_XENTR	Q599j8 xenopus tro
9	1211.5	71.1	304	2 Q5CA57_CHICK	O5ca57 gallus gall
10	1194.5	70.1	330	2 Q599J7_XENLA	Q599j7 xenopus lae
11	1154.5	67.7	304	2 Q5CA27_FUGRU	O5ca27 fugu rubrip
12	1121.5	65.8	304	2 Q5CA28_TETNG	O5ca28 tetraodon n
13	1102.5	64.7	316	2 Q6DBR2_BRARE	O6db2 brachydanio
14	1057.5	62.0	413	2 Q4SG61_TETNG	O4sg61 tetraodon n
15	1055	61.9	303	2 Q599J9_ORYLA	O599j9 oryzias lat
16	909.5	53.3	327	2 Q5CB07_BRARE	Q5cb07 brachydanio
17	787.5	46.2	334	2 Q96FS7_HUMAN	O96fs7 homo sapien
18	787	46.2	335	2 Q5CA29_TETNG	O5ca29 tetraodon n
19	783.5	46.0	334	2 Q5CB02_BRARE	O5cb02 brachydanio
20	779.5	45.7	334	1 B3GA1_HUMAN	O9p2w7 homo sapien
21	779	45.7	335	2 Q5CB01_ORYLA	O5cb01 oryzias lat
22	777.5	45.6	334	1 B3GA1_RAT	O35789 r galactosy
23	777.5	45.6	334	2 Q6PIG8_MOUSE	O6pig8 mus musculu
24	777.5	45.6	347	2 Q8R531_MOUSE	Q8r531 mus musculu
25	777	45.6	335	2 Q5CB00_FUGRU	O5cb00 fugu rubrip
26	776	45.5	335	1 B3GA1_CANFA	O5cb03 c galactosy
27	776	45.5	359	2 Q4SEI4_TETNG	O4sei4 tetraodon n
28	772	45.3	394	2 Q5DTF9_MOUSE	O5d9 mus musculu
29	770.5	45.2	334	1 B3GA1_MOUSE	O9cwt3 mus musculu
30	764	44.8	277	2 Q8RIV2_MOUSE	Q8riv2 mus musculu
31	760	44.6	345	2 Q8BLT8_MOUSE	Q8blt8 mus musculu

Q599k2 bos taurus  
O5cb04 p galactosy  
Q7t180 gallus gall  
O5cb05 fugu rubrip  
O63zm6 xenopus lae  
Q480u6 tetraodon n  
Q5cb06 oryzias lat  
Q599k3 xenopus tro  
O567g3 brachydanio  
Q7t1e2 gallus gall  
Q7zzm8 xenopus lae  
Q8jhi9 brachydanio  
Q599k4 xenopus tro  
Q7t1e1 gallus gall

#### ALIGNMENTS

#### RESULT 1

B3GA2\_HUMAN STANDARD: PRT; 323 AA.  
AC Q9NPZ5; Q5JS09; Q9TF38; Q96NK4;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 13-SEP-2005 (Rel. 48, Last annotation update)  
DE Galactosylgalactosylxylosylprotein 3-beta-glucuronosyltransferase 2  
DE (EC 2.4.1.135) (Beta-1,3-glucuronyltransferase 2)  
DE (Glucuronosyltransferase-S) (GlcAT-S) (UDP-glucuronosyltransferase-S)  
DE (GlcAT-D).  
GN Name=B3GAT2; Synonyms=GLCATS, KIAA1963;  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;  
OC Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP NUCLEOTIDE SEQUENCE.  
RX MEDLINE=22410182; PubMed=12522689; DOI=10.1007/s100380200103;  
RA Marcos I., Galan J.J., Borrego S., Antinolo G.;  
RT "Cloning, characterization, and chromosome mapping of the human GlcAT-S gene.";  
RT J. Hum. Genet. 47:677-680(2002).  
RL [2]  
RN NUCLEOTIDE SEQUENCE [LARGE SCALE MRNA].  
RP TISSUE=Brain;  
RX MEDLINE=21842142; PubMed=11853319;  
RA Nagase T., Kikuno R., Ohara O.;  
RT "Prediction of the coding sequences of unidentified human genes. XXII. The complete sequences of 50 new cDNA clones which code for large proteins.";  
RT DNA Res. 8:319-327(2001).  
RL [3]  
RN NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].  
RX MEDLINE=22935763; PubMed=14574404; DOI=10.1038/nature02055;  
RA Mungall A.J., Palmer S.A., Sims S.K., Edwards C.A., Ashurst J.L., Wilming L., Jones M.C., Horton R., Hunt S.E., Scott C.E., Gilbert J.G.R., Clamp M.E., Bethel G., Milne S., Ainscough R., Almeida J.P., Ambrose K.D., Andrews T.D., Ashwell R.I.S., Babage A.K., Baguley C.L., Bailey J., Banerjee R., Barker D.J., Barlow K.F., Bates K., Beare D.M., Beasley H., Beasley O., Bird C.P., Blakey S.E., Bray-Allen S., Brook J., Brown A.J., Brown J.Y., Burford D.C., Burrill W., Burton J., Carder C., Carter N.P., Chapman J.C., Clark S.Y., Clark G., Clee C.M., Clegg S., Cobley V., Collier R.E., Collins J.E., Colman L.K., Corby N.R., Coville G.J., Cullley K.M., Dhali P., Davies J., Dunn M., Earthrowl M.E., Ellington A.E., Evans K.A., Faulkner L., Francis M.D., Frankish A., Frankland J., French L., Garner P., Garnett J., Ghori M.J., Gilby L.M., Gillson C.J., Glithero R.J., Graham D.V., Grant M., Gribble S., Griffiths C., Griffiths M.N.D., Hall R., Halls K.S., Hammond S., Harley J.D., Hart E.A., Heath P.D., Heathcote R., Holmes S.J., Howden P.J., Howe K.L., Howell G.R., Huckle E., Humphray S.J., Humphries M.D., Hunt A.R., Johnson C.M., Joy A.A.,

RA Kay M., Keenan S.J., Kimberley A.M., King A., Laird G.K., Langford C.,  
RA Lawlor S., Leongamornlert D.A., Leversta M., Lloyd C.R., Lloyd D.M.,  
RA Loveland J.E., Lovell J., Martin S., Mashreghi-Sohamadi M.,  
RA Maslen G.L., Matthews L., McCann O.T., McLaren S.J., McWay K.,  
RA McMurray A., Moore M.J.F., Mullikin J.C., Niblett D., Nickerson T.,  
RA Novik K.L., Oliver K., Overton-Larty E.K., Parker A., Patel R.,  
RA Pearce A.V., Peck A.I., Phillimore B.J.C.T., Phillips S., Plumb R.W.,  
RA Porter K.M., Ramsey Y., Ranby S.A., Rice C.M., Ross M.T., Searle S.W.,  
RA Sehra H.K., Sheridan E., Skuce C.D., Smith S., Smith M., Spraggon L.,  
RA Squares S.L., Steward C.A., Symamore N., Tamiyn-Hall G., Tester J.,  
RA Theaker A.J., Thomas D.W., Thorpe A., Tracey A., Tromans A., Tubby B.,  
RA Wall M., Wallis J.M., West A.P., White S.S., Whitehead S.L.,  
RA Whitaker H., Wild A., Willey D.J., Wilmer T.E., Wood J.M., Wray P.W.,  
RA Wyatt J.C., Young L., Younger R.M., Bentley D.R., Coulson A.,  
RA Durbin R., Hubbard T., Sulston J.E., Dunham I., Rogers J., Beck S.;  
RT "The DNA sequence and analysis of human chromosome 6";  
RL Nature 425:805-811(2003).  
RN (4)  
RP NUCLEOTIDE SEQUENCE [LARGE SCALE MRNA] OF 157-323.  
RC TISSUE=fetal brain;  
RX PubMed=14702039; DOI=10.1038/ng1285;  
RA Oka T., Suzuki Y., Nishikawa T., Otsuki T., Sugiyama T., Irie R.,  
RA Wakamatsu A., Hayashi K., Sato H., Nagai K., Kimura K., Makita H.,  
RA Sekine M., Obayashi M., Nishi T., Shibahara T., Tanaka T., Ishii S.,  
RA Yamamoto J.-I., Saito K., Kawai Y., Isono Y., Nakamura Y.,  
RA Nagahara K., Murakami K., Yasuda T., Iwayanagi T., Wagatsuma M.,  
RA Shiratori A., Sudo H., Hosoiri T., Kaku Y., Kodaira H., Kondo H.,  
RA Sugawara M., Takahashi M., Kanda K., Yokoi T., Furuya T., Kikkawa E.,  
RA Omura Y., Abe K., Kamihara K., Katsuma T., Sato K., Tanikawa M.,  
RA Yamazaki M., Ninomiya K., Iehibashi T., Yamashita H., Murakawa K.,  
RA Fujimori K., Tanai H., Kimata M., Watanabe M., Hiraoka S., Chiba Y.,  
RA Ishida S., Ono Y., Takiguchi S., Watanabe S., Yosida M., Hotuta T.,  
RA Kusano J., Kanehori K., Takahashi-Fujii A., Hara H., Tanase T.-O.,  
RA Nomura Y., Togiya S., Konai F., Hara R., Takeuchi K., Arita M.,  
RA Imose N., Musashino K., Yuuki H., Oshima A., Sasaki N., Aotsuka S.,  
RA Yoshikawa Y., Matsunawa H., Ichihara T., Shiohara T., Sano S.,  
RA Moriya S., Momiyama H., Satoh N., Takami S., Terashima Y., Suzuki O.,  
RA Nakagawa S., Senoh A., Mizoguchi H., Goto Y., Shimizu F., Wakebe H.,  
RA Hishigaki H., Watanabe T., Sugiyama A., Takemoto M., Kawakami B.,  
RA Yamagaki M., Watanabe K., Kumagai A., Itakura S., Fukuzumi Y.,  
RA Fujimori Y., Komiyama M., Tashiro H., Tanigami A., Fujiwara T.,  
RA Ono T., Yamada K., Fujii Y., Ozaki K., Hirao M., Ohmori Y.,  
RA Kawabata A., Hikiji T., Kobatake N., Inagaki H., Ikema Y., Okamoto S.,  
RA Okitani R., Kawakami T., Noguchi H., Goto Y., Shimizu F., Wakebe H.,  
RA Matsumura K., Nakajima Y., Mizuno T., Morinaga M., Sasaki M.,  
RA Togashi T., Oyama M., Hata H., Watanabe M., Komatsu T.,  
RA Mizushima-Sugano J., Satoh T., Shirai Y., Takahashi Y., Nakagawa K.,  
RA Okumura K., Nagase T., Nomura N., Kikuchi H., Masuho Y., Yamashita R.,  
RA Nakai K., Yada T., Nakamura Y., Ohara O., Isogai T., Sugano S.;  
RT "Complete sequencing and characterization of 21,243 full-length human  
cDNAs.";  
RL Nat. Genet. 36:40-45(2004).  
CC -!- FUNCTION: Involved in the biosynthesis of L2/HNK-1 carbohydrate  
CC epitope on both glycolipids and glycoproteins (By similarity).  
CC -!- CATALYTIC ACTIVITY: UDP-glucuronate + 3-beta-D-galactosyl-4-beta-  
CC D-galactosyl-O-beta-D-xylopyranose = UDP + 3-beta-D-  
CC glucuronosyl-3-beta-D-galactosyl-4-beta-D-galactosyl-O-beta-D-  
CC xylopyranose.  
CC -!- COFACTOR: Manganese.  
CC -!- PATHWAY: Glycosylation.  
CC -!- SUBUNIT: Homodimer (Potential).  
CC -!- SUBCELLULAR LOCATION: Type II membrane protein. Golgi.  
CC -!- TISSUE SPECIFICITY: Expressed in the trachea, retina, spinal cord,  
CC hippocampus and other brain regions, and, at lower levels, in  
CC testis and ovary.  
CC -!- SIMILARITY: Belongs to the glycosyltransferase 43 family.  
CC  
CC This Swiss-Prot entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use as long as its content is in no way modified and this statement is not  
CC removed.

DR EMBL: AY070019; AAL57718.1; -; mRNA.  
DR EMBL: AY070110; AAL58977.1; -; Genomic DNA.  
DR EMBL: AY070108; AAL58977.1; JOINED; Genomic DNA.  
DR EMBL: AY070109; AAL58977.1; JOINED; Genomic DNA.  
DR EMBL: AB075843; BAB85549.1; -; ALT INIT; mRNA.  
DR EMBL: AL121961; CA142145.1; -; Genomic DNA.  
DR EMBL: AL450320; CA142145.1; JOINED; Genomic DNA.  
DR EMBL: AL450320; CA139582.1; -; Genomic DNA.  
DR EMBL: AL121961; CA139582.1; JOINED; Genomic DNA.  
DR EMBL: AK055248; BAB70889.1; -; ALT INIT; mRNA.  
DR HSSP: O94766; 1KWS.  
DR Ensembl: ENSG00000112309; Homo sapiens.  
DR HGNC: HGNC:922; B3GAT2.  
DR H-InvDB: HIX00006000; -.  
DR MIM: 607497; -.  
DR GO: GO:0016021; C:integral to membrane; ISS.  
DR GO: GO:0015018; F:galactosylgalactosylxylosylprotein 3-beta-g. .; ISS.  
DR GO: GO:0016757; F:transferase activity, transferring glycosyl. .; ISS.  
DR GO: GO:0016051; P:carbohydrate biosynthesis; ISS.  
DR InterPro: IPR005027; Glyco\_trans\_43.  
DR PANTHER: PTHR10896; Glyco\_trans\_43; 1.  
DR Pfam: PF03360; Glyco\_trans\_43; 1.  
DR Glycoprotein; Golgi stack; Manganese; Metal-binding; Multigene family;  
KW Signal-anchor; Transferase; Transmembrane.  
FT TOPO\_DOM 1 2 Cytoplasmic (Potential).  
FT TRANSMEM 3 23 Signal-anchor for type II membrane protein (Potential).  
FT TOPO\_DOM 24 323 Lumenal (Potential).  
FT CAT\_SITE 273 273 Proton acceptor (By similarity).  
FT METAL 187 187 Manganese (By similarity).  
FT CARBOHYD 67 67 N-linked (GlcNAc. .) (Potential).  
FT CARBOHYD 292 292 N-linked (GlcNAc. .) (Potential).  
SQ SEQUENCE 323 AA; 36919 MW; 85058D52D2D28463 CRC64;  
Query Match 100.0%; Score 1705; DB 1; Length 323;  
Best Local Similarity 100.0%; Pred. No. 2.2e-127; Indels 0; Gaps 0;  
Matches 323; Conservative 0; Mismatches 0;  
QY 1 MKSALFTRFFILLPWILVIMLDVTRRPVPLTPRPYFSPYAVGRCARLPLRRGPA 60  
DB 1 MKSALFTRFFILLPWILVIMLDVTRRPVPLTPRPYFSPYAVGRCARLPLRRGPA 60  
QY 61 HGTQRNSRQPOPEPQLPTIATPTYSRVPQKELTRLANTPROVAQLHWILVEDAA 120  
DB 61 HGTQRNSRQPOPEPQLPTIATPTYSRVPQKELTRLANTPROVAQLHWILVEDAA 120  
QY 121 ARSELVSRFLARAGLPSTHLVPTPRRYKRPGLPRATEORNAGLAWLRHORHQRAQPGV 180  
DB 121 ARSELVSRFLARAGLPSTHLVPTPRRYKRPGLPRATEORNAGLAWLRHORHQRAQPGV 180  
QY 181 LFFADDNDNTYSLELFQEMRTTRKVSVPVGLVGGRRYERPLVENGKVGWYTGWRADRF 240  
DB 181 LFFADDNDNTYSLELFQEMRTTRKVSVPVGLVGGRRYERPLVENGKVGWYTGWRADRF 240  
QY 241 AIDMAGFAVSLQVILSNKPAVKFRGSGQPMQESDFLQITTVBELEPKANNCTKVLVWH 300  
DB 241 AIDMAGFAVSLQVILSNKPAVKFRGSGQPMQESDFLQITTVBELEPKANNCTKVLVWH 300  
QY 301 TRTEKVNLANEPKYHLDVTVKIEV 323  
DB 301 TRTEKVNLANEPKYHLDVTVKIEV 323  
RESULT 2  
Q599K0\_PANTR PRELIMINARY; PRT; 323 AA.  
AC Q599K0;  
DT 10-MAY-2005 (TrEMBLrel. 30, Created)  
DT 10-MAY-2005 (TrEMBLrel. 30, Last sequence update)  
DT 10-MAY-2005 (TrEMBLrel. 30, Last annotation update)  
DE Beta-3-glucuronyltransferase-S (EC 2.4.1.135).  
GN Name=b3gat2;  
OS Pan troglodytes (Chimpanzee).

```

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
OC Pan.
OX NCBI_TaxID=9598;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RA Ouzine M., Magdalou J., Fournel-Gigleux S., Mollicone R., Oriol R.;
RT "Phylogeny of beta3-glucuronosyltransferases.";
RL Submitted (MAR-2005) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ890841; CAI68027.1; -; mRNA.
KW Glycosyltransferase; Transferase.
SQ SEQUENCE 323 AA; 36918 MW; 85058D52D2D28463 CRC64;

Query Match 100.0%; Score 1705; DB 2; Length 323;
Best Local Similarity 100.0%; Pred. No. 2.2e-127;
Matches 323; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MKSALFRFFILLPWILVIIMLDVTRRPVPLTPRPYPSPYAVGGRGRLPLRRGGA 60
Db 1 MKSALFRFFILLPWILVIIMLDVTRRPVPLTPRPYPSPYAVGGRGRLPLRRGGA 60
Qy 61 HGTKRQSRQPQPOPEPOLPTIYAITPTYSRPVQKAEI TRILANTFRQVAQLHWILVEDAA 120
Db 61 HGTKRQSRQPQPOPEPOLPTIYAITPTYSRPVQKAEI TRILANTFRQVAQLHWILVEDAA 120
Qy 121 ARSELVSRFLARAGLPSTHLHVPTRPRYKRPGLPRATEQRNAGLAWLRQRHQRHQAQGV 180
Db 121 ARSELVSRFLARAGLPSTHLHVPTRPRYKRPGLPRATEQRNAGLAWLRQRHQRHQAQGV 180
Qy 181 LFFADDNTYSLELFQEMRTTRKYSVMPVGLVGRRYERPLVGVGWTGWYTGWRDRPF 240
Db 181 LFFADDNTYSLELFQEMRTTRKYSVMPVGLVGRRYERPLVGVGWTGWYTGWRDRPF 240
Qy 241 AIDMAGFAVSLQVILSNPKAVFKRGSGQPGQESDFLKQITTVBELSPKANCTKVLVWH 300
Db 241 AIDMAGFAVSLQVILSNPKAVFKRGSGQPGQESDFLKQITTVBELSPKANCTKVLVWH 300
Qy 301 TRTEKVLNLANEPKYLHLDTVKIEV 323
Db 301 TRTEKVLNLANEPKYLHLDTVKIEV 323

RESULT 3
B3GA2 CANFA
ID B3GA2 CANFA STANDARD; PRT; 329 AA.
AC Q5CAZ6;
DT 10-MAY-2005 (Rel. 47, Created)
DT 10-MAY-2005 (Rel. 47, Last sequence update)
DT 13-SEP-2005 (Rel. 48, Last annotation update)
DE Galactosylgalactosylxylosylprotein 3-beta-glucuronosyltransferase 2
DE (EC 2.4.1.135) (Beta-1,3-glucuronosyltransferase 2)
DE (Glucuronosyltransferase-S) (GlcAT-S) (UDP-glucuronosyltransferase-S)
DE (GlcAT-D).
GN Name=B3GAT2;
OS Canis familiaris (Dog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Laurasiatheria; Carnivora; Fissipedia; Canidae;
OC Canis.
OX NCBI_TaxID=9615;
RN [1]
RP NUCLEOTIDE SEQUENCE (MRNA).
RA Ouzine M., Magdalou J., Fournel-Gigleux S., Mollicone R., Oriol R.;
RT "Phylogeny of beta3-glucuronosyltransferases.";
RL Submitted (MAR-2005) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: Involved in the biosynthesis of L2/HNK-1 carbohydrate epitope on both glycolipids and glycoproteins (By similarity).
CC -!- CATALYTIC ACTIVITY: UDP-glucuronate + 3-beta-D-galactosyl-4-beta-D-galactosyl-1-O-beta-D-xylosylprotein = UDP + 3-beta-D-glucuronosyl-1-O-beta-D-xylosylprotein.
CC -!- COFACTOR: Manganese (By similarity).
CC -!- PATHWAY: Glycosylation.
CC -!- SUBUNIT: Homodimer (Potential).

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
Pan.
NCBI_TaxID=9598;
[1]
NUCLEOTIDE SEQUENCE.
Ouzine M., Magdalou J., Fournel-Gigleux S., Mollicone R., Oriol R.;
"Phylogeny of beta3-glucuronosyltransferases.";
Submitted (MAR-2005) to the EMBL/GenBank/DBJ databases.
EMBL; AJ890841; CAI68027.1; -; mRNA.
Glycosyltransferase; Transferase.
SEQUENCE 323 AA; 36918 MW; 85058D52D2D28463 CRC64;

Query Match 100.0%; Score 1705; DB 2; Length 323;
Best Local Similarity 100.0%; Pred. No. 2.2e-127;
Matches 323; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MKSALFRFFILLPWILVIIMLDVTRRPVPLTPRPYPSPYAVGGRGRLPLRRGGA 60
Db 1 MKSALFRFFILLPWILVIIMLDVTRRPVPLTPRPYPSPYAVGGRGRLPLRRGGA 60
Qy 61 HGTKRQSRQPQPOPEPOLPTIYAITPTYSRPVQKAEI TRILANTFRQVAQLHWILVEDAA 120
Db 61 HGTKRQSRQPQPOPEPOLPTIYAITPTYSRPVQKAEI TRILANTFRQVAQLHWILVEDAA 120
Qy 121 ARSELVSRFLARAGLPSTHLHVPTRPRYKRPGLPRATEQRNAGLAWLRQRHQRHQAQGV 180
Db 121 ARSELVSRFLARAGLPSTHLHVPTRPRYKRPGLPRATEQRNAGLAWLRQRHQRHQAQGV 180
Qy 181 LFFADDNTYSLELFQEMRTTRKYSVMPVGLVGRRYERPLVGVGWTGWYTGWRDRPF 240
Db 181 LFFADDNTYSLELFQEMRTTRKYSVMPVGLVGRRYERPLVGVGWTGWYTGWRDRPF 240
Qy 241 AIDMAGFAVSLQVILSNPKAVFKRGSGQPGQESDFLKQITTVBELSPKANCTKVLVWH 300
Db 241 AIDMAGFAVSLQVILSNPKAVFKRGSGQPGQESDFLKQITTVBELSPKANCTKVLVWH 300
Qy 301 TRTEKVLNLANEPKYLHLDTVKIEV 323
Db 301 TRTEKVLNLANEPKYLHLDTVKIEV 323

RESULT 4
Q599K1 BOVIN
ID Q599K1 BOVIN PRELIMINARY; PRT; 326 AA.
AC Q599K1;
DT 10-MAY-2005 (TrEMBLrel. 30, Created)
DT 10-MAY-2005 (TrEMBLrel. 30, Last sequence update)
DT 10-MAY-2005 (TrEMBLrel. 30, Last annotation update)
DE Beta-3-glucuronosyltransferase-S (EC 2.4.1.135).
GN Name=B3gat2;
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Laurasiatheria; Cetartiodactyla; Ruminantia;
OC Pecora; Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RA Ouzzine M., Magdalou J., Fournel-Gigleux S., Mollicone R., Oriol R.;
RT "Phylogeny of beta3-glucuronosyltransferases.";

```

```

CC -!- SUBCELLULAR LOCATION: Type II membrane protein. Golgi (By
CC similarity).
CC -!- SIMILARITY: Belongs to the glycosyltransferase 43 family.
CC -----
CC This Swiss-Prot entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use as long as its content is in no way modified and this statement is not
CC removed.
CC -----
DR EMBL; AJ888982; CAI62047.1; -; mRNA.
DR InterPro; IPR005027; Glyco_trans_43.
DR PANTHER; PTHR10896; Glyco_trans_43; 1.
DR Pfam; PF03360; Glyco_transf_43; 1.
DR Glycoprotein; Golgi stack; Manganese; Metal-binding; Multigene family;
KW Signal-anchor; Transferase; Transmembrane.
FT TOPO_DOM 1 2 Cytoplasmic (Potential).
FT TRANSMEM 3 23 Signal-anchor for type II membrane
FT protein (Potential).
FT TOPO_DOM 24 329 Luminal (Potential).
FT ACT_SITE 279 279 Proton acceptor (By similarity).
FT METAL 193 193 Manganese (By similarity).
FT CARBOHYD 71 71 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 298 298 N-linked (GlcNAc...) (Potential).
SQ SEQUENCE 329 AA; 37561 MW; 45DF20A242CD1F72 CRC64;

Query Match 95.0%; Score 1620; DB 1; Length 329;
Best Local Similarity 94.5%; Pred. No. 1.3e-120;
Matches 311; Conservative 5; Mismatches 7; Indels 6; Gaps 2;

Qy 1 MKSALFRFFILLPWILVIIMLDVTRRPVPLTPRPYPSPYAVGGRGRLPLRRG--- 57
Db 1 MKSALFRFFILLPWILVIIMLDVTRRPVPLTPRPYPSPYAVGGRGRLPLRRGPD 60
Qy 58 -GPAHTQKQNRSP--QPOPEPOLPTIYAITPTYSRPVQKAEI TRILANTFRQVAQLHWI 114
Db 61 SGFGRGKEKNESEPHARPEPELPTIYAITPTYSRPVQKAEI TRILANTFRQVAQLHWI 120
Qy 115 LVEDAAARSELVSRFLARAGLPSTHLHVPTRPRYKRPGLPRATEQRNAGLAWLRQRHQAQ 174
Db 121 LVEDAAARSELVSRFLARAGLPSTHLHVPTRPRYKRPGLPRATEQRNAGLAWLRQRHQAQ 180
Qy 175 RAQPGVLFFADDNTYSLELFQEMRTTRKYSVMPVGLVGRRYERPLVGVGWTGW 234
Db 181 RAQPGVLFFADDNTYSLELFQEMRTTRKYSVMPVGLVGRRYERPLVGVGWTGW 240
Qy 235 RADRPFAIDMAGFAVSLQVILSNPKAVFKRGSGQPGQESDFLKQITTVBELSPKANCT 294
Db 241 RADRPFAIDMAGFAVSLQVILSNPKAVFKRGSGQPGQESDFLKQITTVBELSPKANCT 300
Qy 295 KVLVWHTRTEKVLNLANEPKYLHLDTVKIEV 323
Db 301 KVLVWHTRTEKVLNLANEPKYLHLDTVKIEV 329

RESULT 4
Q599K1 BOVIN
ID Q599K1 BOVIN PRELIMINARY; PRT; 326 AA.
AC Q599K1;
DT 10-MAY-2005 (TrEMBLrel. 30, Created)
DT 10-MAY-2005 (TrEMBLrel. 30, Last sequence update)
DT 10-MAY-2005 (TrEMBLrel. 30, Last annotation update)
DE Beta-3-glucuronosyltransferase-S (EC 2.4.1.135).
GN Name=B3gat2;
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Laurasiatheria; Cetartiodactyla; Ruminantia;
OC Pecora; Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RA Ouzzine M., Magdalou J., Fournel-Gigleux S., Mollicone R., Oriol R.;
RT "Phylogeny of beta3-glucuronosyltransferases.";

```

```

RL Submitted (MAR-2005) to the EMBL/GenBank/DBJ databases.
DR EMBL: AJ990840; CAL68026.1; -, mRNA.
KW Glycosyltransferase; Transferase.
SQ SEQUENCE 326 AA; 37145 MW; 79ABC47A6177FA73 CRC64;

Query Match          93.4%; Score 1592.5; DB 2; Length 326;
Best Local Similarity 92.6%; Pred. No 2e-118;
Matches 302; Conservative 9; Mismatches 12; Indels 3; Gaps 1;

QY 1 MKSALFTRFFILLPWILVILMLDVTDRPVPPLTPRPVPSYAVGRRGGLPLRRGGPA 60
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1 MKSVLSRPPMLLPWILVILMLDVTDRPVPPLTPRPVPSYAVGRRGGLPLRRGGPG 60
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 61 HGTOKNRSPQ---PQEPQLPTIYAITPTYSRPVQKAEIPLANTFQVQALHWILVE 117
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 61 RGRPRENSWPRPPPPPEPPLTIIYAITPTYSRPVQKAEIPLANTFQVQALHWILVE 120
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 118 DAARSELVSRFLARAGLPSHLVTPTRPKRPGLPRAETEORNAGLWLRQRHQHQAQ 177
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 121 DAARSELVSRFLARAGLPSHLVTPTRPKRPGLPRAETEORNAGLWLRQRHQHQAQ 180
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 178 PGVLFADDDNTYSLELFOEMRTTRKVSVMVPGVLGVRGRYERELVGVGVGMYTGWRA 237
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 181 PGVLFADDDNTYSLELFOEMRTTRKVSVMVPGVLGVRGRYERELVGVGVGMYTGWRA 240
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 238 RPPAIDMAGFAVSLQVILSNPKAVFKRSGSQPMQESDFLKQITVVEELPEPKANNTKVL 297
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 241 RPPAIDMAGFAVSLQVILSNPKAVFKRSGSQPMQESDFLRQITVVEELPEPKANNTKVL 300
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 298 VHTRTKVNLANEPKVLHDTVKIEV 323
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 301 VHTRTKVNLANEPKVLHDTVIDV 326
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 5
B3GA2_MOUSE
ID B3GA2_MOUSE STANDARD; PRT; 324 AA.
AC PS9270;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 13-SEP-2005 (Rel. 48, Last annotation update)
DE Galactosylgalactosylxylosylprotein 3-beta-glucuronosyltransferase 2
DE (EC 2.4.1.135) (Beta-1,3-glucuronyltransferase 2)
DE (Glucuronosyltransferase-S) (GlcAT-S) (UDP-glucuronosyltransferase-S)
DE (GlcAT-D).
DE Names=B3gat2; Synonyms=Glcats;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
OC Muridae; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RN NUCLEOTIDE SEQUENCE (ISOFORM 1).
RC STRAIN=BALB/c; TISSUE=Brain;
RX MEDLINE=22271538; PubMed=12383500; DOI=10.1016/S0378-1119(02)00840-5;
RA Imaiya K., Ishizaki T., Seiki T., Saito F., Inazawa J., Oka S.,
RA Kawasaki T.;
RT "cDNA cloning, genomic structure and chromosomal mapping of the mouse
RT glucuronyltransferase-S involved in the biosynthesis of the HNK-1
RT carbohydrate epitope.";
RL Gene 296:29-36(2002).

[2]
RN NUCLEOTIDE SEQUENCE [LARGE SCALE MRNA] (ISOFORMS 1 AND 2).
RC STRAIN=C57BL/6J; TISSUE=Embryonic head, and Kidney;
RX MEDLINE=22354683; PubMed=12466851; DOI=10.1038/nature01266;
RA Okazaki Y., Furuno M., Kasukawa T., Adachi H., Suzuki H., Yamanaka I., Kiyosawa H.,
RA Nikaido I., Osato N., Saito R., Suzuki H., Bono H., Kondo S.,
RA Yagi K., Tomaru Y., Hasegawa Y., Nogami A., Schonbach C., Gotojori T.,
RA Baldarelli R., Hill D.P., Bult C., Hume D.A., Quackenbush J.,
RA Schriml L.M., Kanapin A., Matsuda H., Batalov S., Beisel K.W.,
RA Blake J.A., Bradt D., Brusci V., Chothia C., Corbani L.E., Cousins S.,
RA Dalla E., Dragani T.A., Fletcher C.F., Forrest A., Frazer K.S.,
RA Gaasterland T., Gariboldi M., Glasi C., Godzik A., Gough J.,

```



```
DR EMBL; AB055902; BAC20343.1; -; mRNA.
DR EMBL; AK048146; BAC33257.1; -; mRNA.
DR EMBL; AK052640; BAC35075.1; -; mRNA.
DR EMBL; BC056368; AAH56368.1; -; mRNA.
DR EMBL; BC058082; AAH58082.1; -; mRNA.
DR HSSP; O94766; 1KWS.
DR Ensemble; ENSMUSG00000026156; Mus musculus.
DR MGI; MGI:2389490; B3gat2.
DR GO; GO:0016021; C:integral to membrane; ISS.
DR GO; GO:0015018; F:galactosylgalactosylxylosylprotein 3-beta-g. . .; ISS.
DR GO; GO:0016757; F:transferase activity, transferring glycosyl. . .; ISS.
DR GO; GO:0016051; F:carbohydrate biosynthesis; ISS.
DR InterPro; IPR005027; Glyco trans 43.
DR PANTHER; PTHR10896; Glyco trans 43; 1.
DR Pfam; PF03360; Glyco trans 43; 1.
KW Alternative splicing; Glycoprotein; Golgi stack; Manganese;
KW Metal-binding; Multigene family; Signal-anchor; Transferrase;
KW Transmembrane.
FT TOPO_DOM 1 2 Cytoplasmic (Potential).
FT TRANSMEM 3 23 Signal-anchor for type II membrane protein (Potential).
FT FT 24 324 Lumenal (Potential).
FT ACT_SITE 24 324 Proton acceptor (By similarity).
FT METAL 188 188 Manganese (by similarity).
FT CARBOHYD 68 68 N-linked (GlcNAc...) (potential).
FT CARBOHYD 293 293 N-linked (GlcNAc...) (potential).
FT VARSPPLIC 248 302 FAVSLQVILSNPKAVFKRRGSGPQMGQSDFLKQITTVVEELE
FT PKASNCTKVLVWHT -> EQNWDICPRMGSPRENWERPIF
FT IKPLSVLHSSSEILIFKIGLSLMVQVEDFRSG (in isoform 2).
FT FT 303 324 /FTID=VSP_001796.
FT VARSPPLIC 303 324 Missing (in isoform 2).
FT CONFLICT 291 291 A -> V (in Ref. 1).
FT SEQUENCE 324 AA; 37132 MW; 675C549D3787B957 CRC64;

Query Match 89.8%; Score 1531.5; DB 1; Length 324;
Best Local Similarity 90.4%; Pred. No. 1.4e-113;
Matches 293; Conservative 10; Mismatches 20; Indels 1; Gaps 1;

Qy 1 MKSALFTRFFILLPWILVIMLDVTRRPVPLTPRPYSPYAVGGRGRLPLRRGGPA 60
Db 1 MKSALCSRRFFLLPWILVIMLDVDRPPAPQITSRPYSPYAVGGRGRLPLRRSSPG 60
Qy 61 H-GTQKNQSRPOPEPOLPTIYAITPTYSRPVQKAELETLANTFRQVQLHILVEDA 119
Db 61 RDAEKKNESRPQLQPEPRLEPTIYAITPTYSRPVQKAELETLANTFRQVQLHILVEDR 120
Qy 120 AARSELYSRFLARAGLPSTHLHVPTPRRYKRPGLPRATEQKNAGLWLRQRHQHQAQPG 179
Db 121 ATRSELVSSFLARAGLPNTHLVPTPRRYKRPWLPATEQKNAGLWLRQRHQHQAQPG 180
Qy 180 VLFADDNTYSLELFQEMRTTRKVSVPVGLVGGRRYRPLVNGKVGWYTTGWADRP 239
Db 181 VLFADDNTYSLELFQEMRTTRKVSVPVGLVGGRRYRPLVNGKVGWYTTGWEDRP 240
Qy 240 FAIDMAGFVSLQVILSNPKAVFKRRGSGPQMGQSDFLKQITTVVEELEPKANCTKVLVW 299
Db 241 FAIDMAGFVSLQVILSNPKAVFKRRGSGPQMGQSDFLKQITTVVEELEPKASNCTKVLVW 300
Qy 300 HTRTEKVNLANEPKYHLDTVKIEV 323
Db 301 HTRTEKVNLANEPKYHLDTVNIEV 324

RESULT 6
Q5DTT7 MOUSE PRELIMINARY; PRT; 349 AA.
AC Q5DTT7,
DT 10-MAY-2005 (TrEMBLrel. 30, Created)
DT 10-MAY-2005 (TrEMBLrel. 30, Last sequence update)
DT 10-MAY-2005 (TrEMBLrel. 30, Last annotation update)
DE MKIAA1963 protein (Fragment).

Name=B3gat2; Synonyms=mkIAA1963;
Mus musculus (Mouse).
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
Muridae; Murinae; Mus.
NCBI_TaxID=10090;
[1]
NUCLEOTIDE SEQUENCE.
TISSUE=Fetal brain;
Okazaki N., Kikuno R.F., Ohara R., Inamoto S., Nagase T., Ohara O.,
Koga H.;
"Prediction of the Coding Sequences of Mouse Homologues of KIAA Gene.
The Complete Nucleotide Sequences of Mouse KIAA-homologous cDNAs
Identified by Screening of Terminal sequences of cDNA Clones Randomly
Sampled from Size-Fractionated Libraries. ";
Submitted (JEB-2005) to the EMBL/GenBank/DBJ databases.
EMBL; AK220433; BAD90265.1; -; mRNA.
MGI; MGI:2389490; B3gat2.
GO; GO:0015020; F:glucuronosyltransferase activity; TAS.
InterPro; IPR005027; Glyco trans 43.
Pfam; PF03360; Glyco trans 43; 1.
NON_TER 1
SEQUENCE 349 AA; 40066 MW; 2EFB2B8BCBD84169 CRC64;

Query Match 89.8%; Score 1531.5; DB 2; Length 349;
Best Local Similarity 90.4%; Pred. No. 1.6e-113;
Matches 293; Conservative 10; Mismatches 20; Indels 1; Gaps 1;

Qy 1 MKSALFTRFFILLPWILVIMLDVTRRPVPLTPRPYSPYAVGGRGRLPLRRGGPA 60
Db 26 MKSALCSRRFFILLPWILVIMLDVDRPPAPQITSRPYSPYAVGGRGRLPLRRSSPG 85
Qy 61 H-GTQKNQSRPOPEPOLPTIYAITPTYSRPVQKAELETLANTFRQVQLHILVEDA 119
Db 86 RDAEKKNESRPQLQPEPRLEPTIYAITPTYSRPVQKAELETLANTFRQVQLHILVEDR 145
Qy 120 AARSELYSRFLARAGLPSTHLHVPTPRRYKRPGLPRATEQKNAGLWLRQRHQHQAQPG 179
Db 146 ATRSELVSSFLARAGLPNTHLVPTPRRYKRPWLPATEQKNAGLWLRQRHQHQAQPG 205
Qy 180 VLFADDNTYSLELFQEMRTTRKVSVPVGLVGGRRYRPLVNGKVGWYTTGWADRP 239
Db 206 VLFADDNTYSLELFQEMRTTRKVSVPVGLVGGRRYRPLVNGKVGWYTTGWEDRP 265
Qy 240 FAIDMAGFVSLQVILSNPKAVFKRRGSGPQMGQSDFLKQITTVVEELEPKANCTKVLVW 299
Db 266 FAIDMAGFVSLQVILSNPKAVFKRRGSGPQMGQSDFLKQITTVVEELEPKASNCTKVLVW 325
Qy 300 HTRTEKVNLANEPKYHLDTVKIEV 323
Db 326 HTRTEKVNLANEPKYHLDTVNIEV 349

RESULT 7
B3GA2 RAT
ID B3GA2 RAT STANDARD; PRT; 324 AA.
AC Q9ZL37,
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 13-SEP-2005 (Rel. 48, Last annotation update)
DE Galactosylgalactosylxylosylprotein 3-beta-glucuronosyltransferase 2
DE (EC 2.4.1.135) (Beta-1,3-glucuronyltransferase 2)
DE (Glucuronosyltransferase-S) (GLCAT-S) (UDP-glucuronosyltransferase-S)
DE (GLCAT-D).
GN Name=B3gat2; Synonyms=Glcats;
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
OC Muroidae; Muridae; Murinae; Rattus.
NCBI_TaxID=10116;
[1]
NUCLEOTIDE SEQUENCE.
TISSUE=Brain;
```

RX MEDLINE=99185317; PubMed=10082676; DOI=10.1006/bbrc.1999.0151;  
 RA Seiki T., Oka S., Terayama K., Imiya K., Kawasaki T.;  
 RT "Molecular cloning and expression of a second glucuronyltransferase  
 involved in the biosynthesis of the HNK-1 carbohydrate epitope.";  
 RL Biochem. Biophys. Res. Commun. 255:182-187(1999).  
 RN [2]  
 RP NUCLEOTIDE SEQUENCE, AND CHARACTERIZATION.  
 RC TISSUE=Embryonic brain;  
 RX MEDLINE=99287914; PubMed=10358066; DOI=10.1074/jbc.274.24.17115;  
 RA Shimoda Y., Tajima Y., Nagase T., Harii K., Osumi N., Sanai Y.;  
 RT "Cloning and expression of a novel galactoside beta1,3-  
 glucuronyltransferase involved in the biosynthesis of HNK-1 epitope.";  
 RL J. Biol. Chem. 274:17115-17122(1999).  
 CC -1- FUNCTION: Involved in the biosynthesis of L2/HNK-1 carbohydrate  
 epitope on both glycolipids and glycoproteins. Substrates include  
 CC asialo-orosomucoid (ASOR), paragloboside (lacto-N-  
 CC neotetraosylceramide), Gal-beta-1,4-GlcNAc-beta-1,3-Gal-beta-1,4-  
 CC Glc-pyridylamine and Gal-beta-1,3-GlcNAc-beta-1,3-Gal-beta-1,4-  
 CC Glc-pyridylamine.  
 CC -1- CATALYTIC ACTIVITY: UDP-glucuronate + 3-beta-D-galactosyl-4-beta-  
 CC D-galactosyl-O-beta-D-xylosylprotein = UDP + 3-beta-D-  
 CC glucuronosyl-3-beta-D-galactosyl-4-beta-D-galactosyl-O-beta-D-  
 CC xylosylprotein.  
 CC -1- COFACTOR: Manganese.  
 CC -1- PATHWAY: Glycosylation.  
 CC -1- SUBUNIT: Homodimer (Potential).  
 CC -1- SUBCELLULAR LOCATION: Type II membrane protein. Golgi.  
 CC -1- TISSUE SPECIFICITY: Expressed in the cerebral cortex, cerebellum  
 CC and whole brain.  
 CC -1- SIMILARITY: Belongs to the glycosyltransferase 43 family.  
 CC  
 CC This Swiss-Prot entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use as long as its content is in no way modified and this statement is not  
 CC removed.  
 CC  
 DR EMBL; AB010441; BAA75219.1; -; mRNA.  
 DR EMBL; AF106624; AAD29576.1; -; mRNA.  
 DR PIR; JG0163; JG0163.  
 DR HSP; O94766; IWS.  
 DR GO; GO:0016021; C:integral to membrane; NAS.  
 DR GO; GO:0015018; F:galactosylgalactosylxylosylprotein 3-beta-g. . . ; IDA.  
 DR GO; GO:0016757; P:transferase activity, transferring glycosyl. . . ; IEP.  
 DR GO; GO:0016051; P:carbohydrate biosynthesis; IEP.  
 DR InterPro; IPR005027; Glyco trans 43.  
 DR PANTHER; PTHR10896; Glyco trans 43; 1.  
 DR Pfam; PF03360; Glyco trans 43; 1.  
 DR Glycoprotein; Golgi stack; Manganese; Metal-binding; Multigene family;  
 KW Signal-anchor; Transferrase; Transmembrane.  
 FT TOPO\_DOM 1 2  
 FT TRANSMEM 3 23  
 FT  
 FT TOPO\_DOM 24 324  
 FT ACT\_SITE 274 274  
 FT METAL 188 188  
 FT CARBOHYD 68 68  
 FT CARBOHYD 293 293  
 FT CARBOHYD 324 AA; 37220 MW; 6283C29535B5D00E CRC64;  
 SQ SEQUENCE 324 AA; 37220 MW; 6283C29535B5D00E CRC64;  
 Query Match 89.5%; Score 1526.5; DB 1; Length 324;  
 Best Local Similarity 89.8%; Pred. No. 3.6e-113;  
 Matches 291; Conservative 10; Mismatches 22; Indels 1; Gaps 1;  
 QY 1 MKSALFTRFPILLPWLIVIMLDVDRPPLTPRPYFSPYAVGGRGRLPLRRGGPA 60  
 Db  
 QY 61 H-GTKRNSRPPQPEPPLPTTIAITPTYSRPVQKAELTRANTFRQAQLHWLVEDA 119  
 Db  
 QY 61 RDAAEKRNESRPPQPEPPLPTTIAITPTYSRPVQKAELTRANTFRQAQLHWLVEDA 120  
 QY 120 AARSELVSRFLARAGLPSTHLHVPTRRYKRPGLPRATEQBNAGLAWLQRHQHQAQPG 179

Db 121 ATRSELVSRFLARAGLPSTHLHVPTRRYKRPGLPRATEQBNAGLAWLQRHQHQAQPG 180  
 QY 180 VLFFADDDNTTYSLELFQEMRTTRKVSVMPLVGLVGGRRYERPLVGVGWTGWRADRP 239  
 Db 181 VLFFADDDNTTYSLELFQEMRTTRKVSVMPLVGLVGGRRYERPLVGVGWTGWRADRP 240  
 QY 240 PAIDNAGFAVSLQVILSNPKAVFKRGSQPGMQESDFLKQITTTVELEPKANNCTKVLVW 299  
 Db 241 PAIDNAGFAVSLQVILSNPKAVFKRGSQPGMQESDFLKQITTTVELEPKANNCTKVLVW 300  
 QY 300 HTRTEKVNLANEPKYHLDVTKIEV 323  
 Db 301 HTRTEKVNLANEPKYHMDVTNIEV 324  
 RESULT 8  
 Q599J8\_XENTR PRELIMINARY; PRT; 331 AA.  
 ID Q599J8\_XENTR PRELIMINARY; PRT; 331 AA.  
 AC Q599J8;  
 DT 10-MAY-2005 (TRENBLrel. 30, Created)  
 DT 10-MAY-2005 (TRENBLrel. 30, Last sequence update)  
 DT 10-MAY-2005 (TRENBLrel. 30, Last annotation update)  
 DE Beta-3-glucuronyltransferase-S (EC 2.4.1.135).  
 GN Name=b3gat2;  
 OS Xenopus tropicalis (Western clawed frog) (Silurana tropicalis).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae;  
 OC Xenopodinae; Xenopus; Silurana.  
 OC NCBI\_TaxID=8364;  
 RN [1]  
 RP NUCLEOTIDE SEQUENCE.  
 RA Quzine M., Magdalou J., Fournel-Gigleux S., Mollicone R., Oriol R.;  
 RT "Phylogeny of beta3-glucuronyltransferases.";  
 RL Submitted (MAR-2005) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AJ789084; CAI68029.1; -; mRNA.  
 KW Glycosyltransferase; Transferrase.  
 SQ SEQUENCE 331 AA; 37869 MW; 21F64C20B8F239D8 CRC64;  
 Query Match 71.6%; Score 1221; DB 2; Length 331;  
 Best Local Similarity 74.1%; Pred. No. 7.6e-89;  
 Matches 246; Conservative 26; Mismatches 50; Indels 10; Gaps 5;  
 QY 1 MKSALFTRFPILLPWLIVIMLDVDRPPLTPRPYFSPYAVGGRGRLPLRRGGP 59  
 Db 1 MKSVFYSRFPILLPWLIVIMLDVDRPPLTPRPYFSPYAVGGRGRLPLRRGGP 59  
 QY 60 AH-GTKRNSRPPQPEPPLPTTIAITPTYSRPVQKAELTRANTFRQAQLHWLVED 118  
 Db 60 SDPSTHQELLSATQRKNETVPIITPTYSRPVQKAELTRANTFRQAQLHWLVED 119  
 QY 119 AARSELVSRFLARAGLPSTHLHVPTRRYKRPGLPRATEQBNAGLAWLQRHQH---OR 175  
 Db 120 SVHTELVSRFLAGAGVTSTHLYVTPRRYKRTGLPRATEQBNAGLAWLQRQPLRT 179  
 QY 176 AQP-----GVLPFADDNTTYSLELFQEMRTTRKVSVMPLVGLVGGRRYERPLVGVGWTG 231  
 Db 180 AQPQDPTGVVFPFADDNTTYSLELFQEMRTTRKVSVMPLVGLVGGRRYERPLVGVGWTG 239  
 QY 232 TCWRADRPFAIDMAGFAVSLQVILSNPKAVFKRGSQPGMQESDFLKQITTTVELEPKAN 291  
 Db 240 TCWRADRPFAIDMAGFAVSLQVILSNPKAVFKRGSQPGMQESDFLKQITTTVELEPKAN 299  
 QY 292 NCTKVLVWHTRTEKVNLANEPKYHLDVTKIEV 323  
 Db 300 NCTKVLVWHTRTEKVNLANEPKYHMDVTNIEV 324  
 RESULT 9  
 Q5CAS7\_CHICK  
 ID Q5CAS7\_CHICK PRELIMINARY; PRT; 304 AA.  
 AC Q5CAS7;  
 DT 10-MAY-2005 (TRENBLrel. 30, Created)





Search completed: May 9, 2006, 10:55:21  
Job time : 112 secs

**This Page Blank (uspto)**

GenCore version 5.1.8  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: May 9, 2006, 10:51:05 ; Search time 86 Seconds  
(without alignments)  
1650.225 Million cell updates/sec

Title: US-10-634-905-2

Perfect score: 1705

Sequence: 1 MKSALFTRFFILLPWILIVI.....EKNLANEPKHYLDVTVKIEV 323

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2443163 seqs, 439378781 residues

Total number of hits satisfying chosen parameters: 2443163

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A\_Geneseq\_21:\*

- 1: Geneseq1980s:\*
- 2: Geneseq1990s:\*
- 3: Geneseq2000s:\*
- 4: Geneseq2001s:\*
- 5: Geneseq2002s:\*
- 6: Geneseq2003as:\*
- 7: Geneseq2003bs:\*
- 8: Geneseq2004s:\*
- 9: Geneseq2005s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1705	100.0	323	5	ADI03930 Human enz
2	1705	100.0	323	8	ADQ97589 Human can
3	1526.5	89.5	324	5	ADI03932 Rat galac
4	1223	71.7	292	8	ADQ97586 Mouse can
5	1039	60.9	197	5	ADI03933 HumanUDP-
6	787.5	46.2	347	5	ABBB97603 Novel hum
7	779.5	45.7	334	6	ABO07264 Human p53
8	777.5	45.6	347	2	RAY08214 Rat Glucu
9	777.5	45.6	347	2	AAW90057 Rat GicAT
10	653	38.3	335	6	ABO07265 Human p53
11	653	38.3	335	8	ADO31194 Human GT
12	653	38.3	335	8	ADQ21418 Human sof
13	645	37.8	335	9	ADY70386 Galactosy
14	608	35.7	305	8	ADI00832 Fruit fly
15	510	29.9	241	4	AAW93879 Human pol
16	510	29.9	241	8	ADL31964 Human pro
17	502.5	29.5	316	8	ADI00836 Fruit fly
18	502.5	29.5	479	4	ABB62172 Drosophil
19	487.5	28.6	366	8	ADI00834 Fruit fly
20	487.5	28.6	443	4	ABB60238 Drosophil
21	422	24.8	145	4	AAW99900 Human exc
22	422	24.8	145	4	AAW43699 Human bla
23	422	24.8	145	8	ADF71658 Human bla
24	406	23.8	691	4	ABG00332 Novel hum

25	324.5	19.0	201	8	ADO57267	Ado57267 Kidney de
26	305.5	17.9	308	3	AAG36689	Aag36689 Arabidops
27	305.5	17.9	317	3	AAG36688	Aag36688 Arabidops
28	272.5	16.0	241	3	AAG36690	Aag36690 Arabidops
29	262	15.4	59	3	AAB45114	Aab45114 Human sec
30	246	14.4	479	8	ADY13062	Ady13062 Plant ful
31	241.5	14.2	460	8	ADY06832	Ady06832 Plant ful
32	227	13.3	411	8	ADX91479	Adx91479 Plant ful
33	209.5	12.3	316	8	ADX90903	Adx90903 Plant ful
34	194.5	11.4	167	4	ABG13553	Abg13553 Novel hum
35	193.5	11.3	492	5	ABP99383	Abp99383 Arabidops
36	193.5	11.3	492	5	ABP99383	Abp99383 Arabidops
37	180	10.6	59	3	AAB45113	Aab45113 Herbicida
38	171	10.0	273	8	ADX87904	Adx87904 Plant ful
39	169	9.9	544	5	ABB93266	Abb93266 Herbicida
40	169	9.9	544	5	ABB93284	Abb93284 Herbicida
41	150	8.8	29	3	AAB45116	Aab45116 Human sec
42	141.5	8.3	378	8	ADX72765	Adx72765 Plant ful
43	139	8.2	40	8	ABO59511	AbO59511 Human gen
44	137.5	8.1	267	8	ADX91941	Adx91941 Plant ful
45	130.5	7.7	521	7	ABO69373	AbO69373 Pseudomon

ALIGNMENTS

RESULT 1

ADI03930

ID ADI03930 standard; protein; 323 AA.

XX AC ADI03930;

XX AC

DT 22-APR-2004 (first entry)

XX Human enzyme protein.

DE Human enzyme protein.

XX Enzyme protein; drug screening; disease diagnosis; human; gene therapy;

XX Enzyme protein; drug screening; disease diagnosis; human; gene therapy;

XX Enzyme protein; drug screening; disease diagnosis; human; gene therapy;

XX Enzyme protein; drug screening; disease diagnosis; human; gene therapy;

XX Enzyme protein; drug screening; disease diagnosis; human; gene therapy;

XX Enzyme protein; drug screening; disease diagnosis; human; gene therapy;

XX Enzyme protein; drug screening; disease diagnosis; human; gene therapy;

XX Enzyme protein; drug screening; disease diagnosis; human; gene therapy;

XX Enzyme protein; drug screening; disease diagnosis; human; gene therapy;

XX Enzyme protein; drug screening; disease diagnosis; human; gene therapy;

XX Enzyme protein; drug screening; disease diagnosis; human; gene therapy;

XX Enzyme protein; drug screening; disease diagnosis; human; gene therapy;

XX Enzyme protein; drug screening; disease diagnosis; human; gene therapy;

XX Enzyme protein; drug screening; disease diagnosis; human; gene therapy;

XX Enzyme protein; drug screening; disease diagnosis; human; gene therapy;

XX Enzyme protein; drug screening; disease diagnosis; human; gene therapy;

XX Enzyme protein; drug screening; disease diagnosis; human; gene therapy;

XX Enzyme protein; drug screening; disease diagnosis; human; gene therapy;

XX Enzyme protein; drug screening; disease diagnosis; human; gene therapy;

XX Enzyme protein; drug screening; disease diagnosis; human; gene therapy;

XX Enzyme protein; drug screening; disease diagnosis; human; gene therapy;

XX Enzyme protein; drug screening; disease diagnosis; human; gene therapy;

XX Enzyme protein; drug screening; disease diagnosis; human; gene therapy;

XX Enzyme protein; drug screening; disease diagnosis; human; gene therapy;

XX Enzyme protein; drug screening; disease diagnosis; human; gene therapy;

XX Enzyme protein; drug screening; disease diagnosis; human; gene therapy;

XX Enzyme protein; drug screening; disease diagnosis; human; gene therapy;

XX Enzyme protein; drug screening; disease diagnosis; human; gene therapy;

XX Enzyme protein; drug screening; disease diagnosis; human; gene therapy;

XX Enzyme protein; drug screening; disease diagnosis; human; gene therapy;

XX Enzyme protein; drug screening; disease diagnosis; human; gene therapy;

XX Enzyme protein; drug screening; disease diagnosis; human; gene therapy;

XX Enzyme protein; drug screening; disease diagnosis; human; gene therapy;

XX Enzyme protein; drug screening; disease diagnosis; human; gene therapy;

XX Enzyme protein; drug screening; disease diagnosis; human; gene therapy;

XX Enzyme protein; drug screening; disease diagnosis; human; gene therapy;

XX Enzyme protein; drug screening; disease diagnosis; human; gene therapy;

XX Enzyme protein; drug screening; disease diagnosis; human; gene therapy;

XX Enzyme protein; drug screening; disease diagnosis; human; gene therapy;

XX Enzyme protein; drug screening; disease diagnosis; human; gene therapy;

XX Enzyme protein; drug screening; disease diagnosis; human; gene therapy;

XX Enzyme protein; drug screening; disease diagnosis; human; gene therapy;

XX Enzyme protein; drug screening; disease diagnosis; human; gene therapy;

CC fluids, or as markers for tissues in which the corresponding protein is  
CC preferentially expressed. The human enzyme protein is also useful for  
CC diagnosing a disease, predisposition to a disease, or treating a disorder  
CC characterized by an absence of, inappropriate or unwanted expression of  
CC the protein. The antibodies are useful in pharmacogenomic analysis, for  
CC inhibiting protein function, or for tissue typing. The nucleic acid  
CC molecules are useful as probes, primers, chemical intermediates, or in  
CC biological assays. The present sequence represents a human enzyme  
CC protein.  
XX  
SQ Sequence 323 AA;

Query Match 100.0%; Score 1705; DB 5; Length 323;  
Best Local Similarity 100.0%; Pred. No. 7.5e-161;  
Matches 323; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 MKSALFTRFFILLPWILIVIMLDVDTTRRPVPLTPRPYFSPYAVGGRGARLPRLRGGA 60  
DB 1 MKSALFTRFFILLPWILIVIMLDVDTTRRPVPLTPRPYFSPYAVGGRGARLPRLRGGA 60  
QY 61 HGTQKRNQSRPQPEPQPLPTIYAITPTYSRPVQKAELTRLANTFRQVQLHWILVEDAA 120  
DB 61 HGTQKRNQSRPQPEPQPLPTIYAITPTYSRPVQKAELTRLANTFRQVQLHWILVEDAA 120  
QY 121 ARSELVSRFLARAGLPSTHLHVPTPRYKRPGLPRATEQORNAGLAWLRORHQRHQAQPGV 180  
DB 121 ARSELVSRFLARAGLPSTHLHVPTPRYKRPGLPRATEQORNAGLAWLRORHQRHQAQPGV 180  
QY 181 LFFADDDNTYSLELFQEMRTTRKVSVPVGLVGGRRYERPLVGVGWYTGWRADRP 240  
DB 181 LFFADDDNTYSLELFQEMRTTRKVSVPVGLVGGRRYERPLVGVGWYTGWRADRP 240  
QY 241 AIDMAGFAVSLQVILSNPKAVFKRRGSGQPMQESDFLKQITTVVELEPKANNCTKVLVWH 300  
DB 241 AIDMAGFAVSLQVILSNPKAVFKRRGSGQPMQESDFLKQITTVVELEPKANNCTKVLVWH 300  
QY 301 TRTEKVNLANEPKYHLDTVKIEV 323  
DB 301 TRTEKVNLANEPKYHLDTVKIEV 323

RESULT 2  
ADQ97589  
ID ADQ97589 standard; protein; 323 AA.  
XX AC ADQ97589;  
XX DT 07-OCT-2004 (first entry)  
XX DE Human cancer associated sequence HP10-009, SEQ ID 566.  
XX KW Cytostatic; Gene Therapy; cancer; leukemia; lymphoma; Human.  
XX OS Homo sapiens.  
XX FN WO2004060304-A2.  
XX PD 22-JUL-2004.  
XX PF 22-DEC-2003; 2003WO-US041389.  
XX PR 27-DEC-2002; 2002US-00330773.  
XX PA (SAGR-) SAGRES DISCOVERY INC.  
XX PI Morris DW, Malandro MS;  
XX DR WPI; 2004-543781/52.  
XX PT New isolated cancer associated nucleic acids comprising at least 10  
XX PT contiguous nucleotides, useful for diagnosing, preventing and/or treating  
XX PT cancers such as leukemia and lymphoma.

PS Claim 1; SEQ ID NO 566; 199pp; English.  
XX The present invention relates to cancer associated sequences (ADQ97025-  
CC ADQ98004). The sequences are useful for the diagnosis, prevention and/or  
CC treatment of cancer, such as leukemia and lymphoma. Note: The sequence  
CC data for this patent did not form part of the printed specification, but  
CC was obtained in electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 323 AA;

Query Match 100.0%; Score 1705; DB 8; Length 323;  
Best Local Similarity 100.0%; Pred. No. 7.5e-161;  
Matches 323; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 MKSALFTRFFILLPWILIVIMLDVDTTRRPVPLTPRPYFSPYAVGGRGARLPRLRGGA 60  
DB 1 MKSALFTRFFILLPWILIVIMLDVDTTRRPVPLTPRPYFSPYAVGGRGARLPRLRGGA 60  
QY 61 HGTQKRNQSRPQPEPQPLPTIYAITPTYSRPVQKAELTRLANTFRQVQLHWILVEDAA 120  
DB 61 HGTQKRNQSRPQPEPQPLPTIYAITPTYSRPVQKAELTRLANTFRQVQLHWILVEDAA 120  
QY 121 ARSELVSRFLARAGLPSTHLHVPTPRYKRPGLPRATEQORNAGLAWLRORHQRHQAQPGV 180  
DB 121 ARSELVSRFLARAGLPSTHLHVPTPRYKRPGLPRATEQORNAGLAWLRORHQRHQAQPGV 180  
QY 181 LFFADDDNTYSLELFQEMRTTRKVSVPVGLVGGRRYERPLVGVGWYTGWRADRP 240  
DB 181 LFFADDDNTYSLELFQEMRTTRKVSVPVGLVGGRRYERPLVGVGWYTGWRADRP 240  
QY 241 AIDMAGFAVSLQVILSNPKAVFKRRGSGQPMQESDFLKQITTVVELEPKANNCTKVLVWH 300  
DB 241 AIDMAGFAVSLQVILSNPKAVFKRRGSGQPMQESDFLKQITTVVELEPKANNCTKVLVWH 300  
QY 301 TRTEKVNLANEPKYHLDTVKIEV 323  
DB 301 TRTEKVNLANEPKYHLDTVKIEV 323

RESULT 3  
ADI03932  
ID ADI03932 standard; protein; 324 AA.  
XX AC ADI03932;  
XX DT 22-APR-2004 (first entry)  
XX DE Rat galactosidase beta-1,3-glucuronyltransferase.  
XX KW Enzyme protein; drug screening; disease diagnosis; rat; gene therapy;  
XX KW chromosome 6; enzyme; glucuronyltransferase;  
XX KW galactosidase beta-1,3-glucuronyltransferase.  
XX OS Rattus norvegicus.  
XX FN WO200268657-A2.  
XX PD 06-SEP-2002.  
XX PF 08-FEB-2002; 2002WO-US003623.  
XX PR 26-FEB-2001; 2001US-0270871P.  
XX PR 26-MAR-2001; 2001US-00816095.  
XX PA (PEXE ) PE CORP NY.  
XX PI Gan W, Yan C, Merkulov GV, Di Francesco V, Beasley EM;  
XX WPI; 2002-713380/77.  
XX DR REFSEQ; NP\_072131.1.  
XX PT New human enzyme proteins, useful for treating or diagnosing disorders



associated with abnormal expression of the protein, in drug screening assays and pharmacogenomic analysis.

Disclosure; SEQ ID NO 4; 127pp; English.

The invention relates to a novel isolated enzyme protein and encoding polynucleotides. The protein shows a high degree of similarity to a glucuronyltransferase cloned from a rabbit brain cDNA library. The peptides and nucleic acid molecules are useful as models for the development of human therapeutic targets, aid in the identification of therapeutic proteins, and serve as targets for the development of human therapeutic agents. The peptide may be used in drug screening assays, in assays to determine the biological activity of the protein, to raise antibodies or to elicit another immune response, as a reagent in assays designed to quantitatively determine levels of the protein in biological fluids, or as markers for tissues in which the corresponding protein is preferentially expressed. The human enzyme protein is also useful for diagnosing a disease, predisposition to a disease, or treating a disorder characterized by an absence of, inappropriate or unwanted expression of the protein. The antibodies are useful in pharmacogenomic analysis, for inhibiting protein function, or for tissue typing. The nucleic acid molecules are useful as probes, primers, chemical intermediates, or in biological assays. The present sequence represents a rat galactosidase beta-1,3-glucuronyltransferase used in alignment studies with the novel human enzyme protein.

Sequence 324 AA;

Query Match 89.5%; Score 1526.5; DB 5; Length 324; Best Local Similarity 89.8%; Pred. No. 4.6e-143; Matches 291; Conservative 10; Mismatches 22; Indels 1; Gaps 1;

QY 1 M K S A L T R F F I L L P W I L I V I I M L D V T R P V P L T R P Y F S P V A V G R G G A R L P L R R G G P A 60  
DB 1 M K S A L C N R F F I L L P W I L I V I I M L D V D P R P A P Q L T S R P Y F S P H A V G C G S R V P L R S S P G 60  
QY 61 H - G T Q K R N Q S R P Q P E P O L P T T I A I T P T Y S R P V Q K A E L T R L A N T F R Q V A Q L H W I L V E D A 119  
DB 61 R D A E K R N E S R P Q L P E P L P T T I A I T P T Y S R P V Q K A E L T R L A N T F R Q V A Q L H W I L V E D R 120  
QY 120 A A R S E L V S R F L A R A G L P S T H L V P T R R Y K R P G L P R A T E Q R N A G L A W L R Q R H Q H Q A Q P G 179  
DB 121 A T R S E L V S S F L A R A G L P N T H L V P T R R Y K R P W L P R A T E Q R N A G L A W L R Q R H Q H Q A Q P G 180  
QY 180 V L F F A D D N T Y S L E L F Q E M T T R K V S V M P V G L V G R R Y E R P L V K N G K V G W Y T G W R A D R P 239  
DB 181 V L F F A D D N T Y S L E L F Q E M T T R K V S V M P V G L V G R R Y E R P L V K N G K V G W Y T G W R E D R P 240  
QY 240 F A I D M A G F A V S L Q V I L S N P K A V F K R G S Q P G M Q E S D F L K Q I T T V E E L E P K A N N C T K V L W 299  
DB 241 F A I D M A G F A V S L Q V I L S N P K A V F K R G S Q P G M Q E S D F L K Q I T T V E E L E P K A N N C T K V L W 300  
QY 300 H T R T E K V N L A N E P K Y H L D T V K I E V 323  
DB 301 H T R T E K V N L A N E P K Y H M D T V N I E V 324

RESULT 4  
ADQ97586  
ID ADQ97586 standard; protein; 292 AA.

AC ADQ97586;  
XX  
XX 07-OCT-2004 (first entry)  
XX  
XX Mouse cancer associated sequence MP10-009, SEQ ID 563.  
XX  
XX Cytostatic; Gene Therapy; cancer; leukemia; lymphoma; Mouse.  
XX  
XX Mus musculus.  
XX  
XX WO2004060304-A2.

PD 22-JUL-2004.  
XX  
XX 22-DEC-2003; 2003WO-US041389.  
XX  
XX 27-DEC-2002; 2002US-00330773.  
XX  
XX (SAGR-) SAGRES DISCOVERY INC.  
XX  
XX Morris DW, Malandro MS;  
XX  
XX WPI; 2004-543781/52.

+

XX  
XX New isolated cancer associated nucleic acids comprising at least 10  
XX contiguous nucleotides, useful for diagnosing, preventing and/or treating  
XX cancers such as leukemia and lymphoma.  
XX  
XX Claim 1; SEQ ID NO 563; 199pp; English.

XX  
XX The present invention relates to cancer associated sequences (ADQ97025-  
XX ADQ98004). The sequences are useful for the diagnosis, prevention and/or  
XX treatment of cancer, such as leukemia and lymphoma. Note: The sequence  
XX data for this patent did not form part of the printed specification, but  
XX was obtained in electronic format directly from WIPO at  
XX ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
XX Sequence 292 AA;

Query Match 71.7%; Score 1223; DB 8; Length 292;  
Best Local Similarity 74.1%; Pred. No. 6.9e-113;  
Matches 240; Conservative 9; Mismatches 17; Indels 58; Gaps 2;

QY 1 M K S A L T R F F I L L P W I L I V I I M L D V T R P V P L T R P Y F S P V A V G R G G A R L P L R R G G P A 60  
DB 26 M K S A L C S R F F I L L P W I L I V I I M L D V D P R P A P Q L T S R P Y F S P H A V G C G S R V P L R S S P G 85  
QY 61 H - G T Q K R N Q S R P Q P E P O L P T T I A I T P T Y S R P V Q K A E L T R L A N T F R Q V A Q L H W I L V E D A 119  
DB 86 R D A E K R N E S R P Q L P E P L P T T I A I T P T Y S R P V Q K A E L T R L A N T F R Q V A Q L H W I L V E D R 118  
QY 120 A A R S E L V S R F L A R A G L P S T H L V P T R R Y K R P G L P R A T E Q R N A G L A W L R Q R H Q H Q A Q P G 179  
DB 119 A T R S E L V S S F L A R A G L P N T H L V P T R R Y K R P W L P R A T E Q R N A G L A W L R Q R H Q H Q A Q P G 148  
QY 180 V L F F A D D N T Y S L E L F Q E M T T R K V S V M P V G L V G R R Y E R P L V K N G K V G W Y T G W R A D R P 239  
DB 149 V L F F A D D N T Y S L E L F Q E M T T R K V S V M P V G L V G R R Y E R P L V K N G K V G W Y T G W R E D R P 208  
QY 240 F A I D M A G F A V S L Q V I L S N P K A V F K R G S Q P G M Q E S D F L K Q I T T V E E L E P K A N N C T K V L W 299  
DB 209 F A I D M A G F A V S L Q V I L S N P K A V F K R G S Q P G M Q E S D F L K Q I T T V E E L E P K A N N C T K V L W 268  
QY 300 H T R T E K V N L A N E P K Y H L D T V K I E V 323  
DB 269 H T R T E K V N L A N E P K Y H M D T V N I E V 292

RESULT 5  
AD103933  
ID AD103933 standard; protein; 197 AA.

XX  
XX AD103933;  
XX  
XX 22-APR-2004 (first entry)  
XX  
XX Human UDP-glucuronyltransferase-S fragment.  
XX  
XX Enzyme protein; drug screening; disease diagnosis; rat; gene therapy;  
XX chromosome 6; enzyme; glucuronyltransferase; UDP-glucuronyltransferase-S.  
XX  
XX Homo sapiens.  
XX  
XX WO200268657-A2.

PD 06-SEP-2002.  
XX  
PF 08-FEB-2002; 2002WO-US003623.  
XX  
XX 26-FEB-2001; 2001US-0270871P.  
PR 26-MAR-2001; 2001US-00816095.  
XX  
XX (PEKE ) PE CORP NY.  
XX  
XX Gan W, Yan C, Merkulov GV, Di Francesco V, Beasley EM;  
XX WPI; 2002-713380/77.  
XX  
XX New human enzyme proteins, useful for treating or diagnosing disorders  
PT associated with abnormal expression of the protein, in drug screening  
PT assays and pharmacogenomic analysis.  
XX  
XX Disclosure; SEQ ID NO 5; 127pp; English.  
XX  
XX The invention relates to a novel isolated enzyme protein and encoding  
CC polynucleotides. The protein shows a high degree of similarity to a  
CC glucuronyltransferase cloned from a rabbit brain cDNA library. The  
CC peptides and nucleic acid molecules are useful as models for the  
CC development of human therapeutic targets, aid in the identification of  
CC therapeutic proteins, and serve as targets for the development of human  
CC therapeutic agents. The peptide may be used in drug screening assays, in  
CC assays to determine the biological activity of the protein, to raise  
CC antibodies or to elicit another immune response, as a reagent in assays  
CC designed to quantitatively determine levels of the protein in biological  
CC fluids, or as markers for tissues in which the corresponding protein is  
CC preferentially expressed. The human enzyme protein is also useful for  
CC diagnosing a disease, predisposition to a disease, or treating a disorder  
CC characterized by an absence of, inappropriate or unwanted expression of  
CC the protein. The antibodies are useful in pharmacogenomic analysis, for  
CC inhibiting protein function, or for tissue typing. The nucleic acid  
CC molecules are useful as probes, primers, chemical intermediates, or in  
CC biological assays. The present sequence represents a human UDP-  
CC glucuronyltransferase-S fragment used in alignment studies with the novel  
CC human enzyme protein.  
XX  
SQ Sequence 197 AA;

Query Match 60.9%; Score 1039; DB 5; Length 197;  
Best Local Similarity 100.0%; Pred. No. 8.5e-95;  
Matches 197; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 MKSALTRFRFILLPWILLIIMLDVTRRPVPLTPRPYFSPYAVGGRGARLPRLRGGA 60  
DB 1 MKSALTRFRFILLPWILLIIMLDVTRRPVPLTPRPYFSPYAVGGRGARLPRLRGGA 60  
QY 61 HGTKQNRQSRPQPEPQLPTTIVAITPTYSRPVQKAELTRLANTFPRQVQLHWILVEDAA 120  
DB 61 HGTKQNRQSRPQPEPQLPTTIVAITPTYSRPVQKAELTRLANTFPRQVQLHWILVEDAA 120  
QY 121 ARSELVSRLARAGLSTHLHVTTPRYPKRGPLPRATEQGNAGLAWLRQHRQHORAPGV 180  
DB 121 ARSELVSRLARAGLSTHLHVTTPRYPKRGPLPRATEQGNAGLAWLRQHRQHORAPGV 180  
QY 181 LFFADDNDNTYSLFQEB 197  
DB 181 LFFADDNDNTYSLFQEB 197

RESULT 6  
ABB97603  
ID ABB97603 standard; protein; 347 AA.  
XX  
AC ABB97603;  
XX  
XX 27-JUN-2002 (first entry)  
DT  
XX Novel human protein SEQ ID NO: 871.  
DE  
XX

KW Human; antianaemic; vulnary; antiinflammatory; immunomodulator;  
KW antinfertility; cerebroprotective; cytotoxic; rheumatic; gene therapy;  
KW neuroprotective; antiparkinsonian; protein therapy; EST;  
XX expressed sequence tag.  
OS Homo sapiens.  
XX WO200222660-A2.  
XX  
XX 21-MAR-2002.  
XX  
XX 10-SEP-2001; 2001WO-US026015.  
XX  
XX 11-SEP-2000; 2000US-00659671.  
XX  
XX (HYSE-) HYSEQ INC.  
XX  
XX Tang YT, Liu C, Zhou P, Asundi V, Zhang J, Zhao QA, Ren F;  
PI Xue AJ, Yang Y, Wehrman T, Drmanac RT;  
XX  
XX WPI; 2002-292408/33.  
DR N-PSDB; ABN32789.  
XX  
XX An isolated polynucleotide for treating diseases associated with its  
PT encoded polypeptide such as cancer and multiple sclerosis.  
XX  
XX Example 2; SEQ ID NO 871; 509pp; English.  
XX  
XX The present invention provides the protein and coding sequences of 444  
CC novel human proteins. These were isolated from expressed sequences tags  
CC (ESTs). They can be used to stimulate cell growth, to regulate  
CC haematopoiesis e.g. to treat aplastic anaemia, to help tissue regrowth  
CC e.g. in burn treatment, to regulate the immune system e.g. to treat  
CC multiple sclerosis, to regulate activin or inhibin e.g. to treat  
CC infertility, to regulate haemostasis or thrombolysis e.g. to treat stroke  
CC and cancer, to screen for drugs, to treat inflammatory conditions e.g.  
CC rheumatoid arthritis, and to treat nervous system disorders e.g.  
CC Parkinson's disease. The present sequence is a protein of the invention  
XX  
SQ Sequence 347 AA;

Query Match 46.2%; Score 787.5; DB 5; Length 347;  
Best Local Similarity 49.0%; Pred. No. 2.2e-69;  
Matches 169; Conservative 37; Mismatches 84; Indels 55; Gaps 8;  
QY 11 ILLPWILIVII-----MLDV-----DTRRPVPP-LTPRPYFSPYAVGGRGARLPRLR 55  
DB 26 IVLPWTLITVWQSTLAPLAVHKDEGSDPRRPTPGADPREYC----- 70  
QY 56 RGGPAHGTQKRN-----QSRPQPEPQLPTTIVAITPTYSRPVQKAELTRLANTF 105  
DB 71 -----TSDRDIVEVWRTEYVYTRPPWSD-TLPTIHVVTPYSRPVQKAELTRMANTL 122  
QY 106 ROVALQHLWILVEDAAARSELVSRLARAGLSTHLHVTTPREYK-----RPLDPRATEQ 159  
DB 123 LHVPLNLWLWVEDAPRRTPLTARLLRDTGLNTHLVETPRNYKLRGDAARDPRIPRGTNQ 182  
QY 160 RNAGLAWLRQHRQHORAPGVLFADDNDNTYSLFQEMRTTRKYSVWPVGLVGRRYER 219  
DB 183 RNIALRWLRETFPRNSSQPGVYVYFADDNDNTYSLFQEMRSTRYVWPVAVFGGLRYEA 242  
QY 220 PLVEN-GKVVGYWTGWRADRPFAIDMAGFAVLSQVILNSPNKAVFKRRSQPMQSDFLK 278  
DB 243 PRVNGAGKVVGWKTVDFPHRPFADIDMAGFAVNLRLILQRSQAYFKLRGVKGGYQESSLLR 302  
QY 279 QITTVEELEPKANCTKVLVWHTRTKYNLANEPKYHLDTVKIEV 323  
DB 303 ELVTINDLEPKAANCTKILVWHTRTKEPKVPLEVNEGKKGTDPDSVEI 347  
RESULT 7  
ABO07264  
ID ABO07264 standard; protein; 334 AA.



[illegible]



XX soft tissue sarcoma; cytostatic; gene therapy; vaccine; screening; human.  
KW Homo sapiens.  
XX WO2004048938-A2.  
XX 10-JUN-2004.  
XX 26-NOV-2003; 2003WO-US038193.  
XX 26-NOV-2002; 2002US-0429739P.  
XX (PROT-) PROTEIN DESIGN LABS INC.  
XX Aziz N, Ginsburg WM, Zlotnik A;  
XX WPI; 2004-441208/41.  
XX Early detection of soft tissue sarcoma comprises determining expression  
PT of a gene in a first soft tissue sample and a normal soft tissue sample  
PT and comparing the gene expression, also useful in treating soft tissue  
PT sarcoma.  
XX Example 2; SEQ ID NO 4238; 210pp; English.  
XX The invention relates to a novel method for detecting soft tissue sarcoma  
CC which comprises obtaining a first soft tissue sample from an individual  
CC and a normal soft tissue sample from the same or different individual,  
CC determining the expression of a gene in both samples and comparing the  
CC expression of the gene in both soft tissue samples, where a higher level  
CC of protein expression in the first soft tissue sample indicates the  
CC presence of soft tissue sarcoma. The method of the invention has  
CC cytotatic applications and may be useful for detecting soft tissue  
CC sarcoma, possibly via gene therapy or vaccine production. The nucleic  
CC acid sequences may be useful in diagnostic and screening applications.  
CC The current sequence is that of a human soft tissue sarcoma-upregulated  
CC protein of the invention. The current sequence is not shown within the  
CC specification per se but was submitted in CD format by the inventor.  
XX SQ Sequence 335 AA;  
Query Match 38.3%; Score 653; DB 8; Length 335;  
Best Local Similarity 43.9%; Pred. No. 5.3e-56;  
Matches 150; Conservative 41; Mismatches 93; Indels 58; Gaps 10;  
QY 1 MKSALFTRFFILLPWILVIMLDVDTTRP---VPPLTPRPYSPYAVGRGGA-----50  
DB 5 LKNVFLAYFLVSIAGLLVALVQLG---QPCDCLPPL-----RAAAEQLRQKD 48  
QY 51 -----RLPLRRGGPAHGTQKRNQSRPQPQPEQ-LPTIYAITPTYSRPVQKAELTRLAN 103  
DB 49 LRISQLQAEULRRPPA-----PAQPEPEALPTIIVVTTVARLVQKAEVLRLSQ 98  
QY 104 TFRQVLAQLHWILVEDAAASSELVSRFLARAGLSTHLVPTPR----RYKRPGL--PRAT 157  
DB 99 TLSLVPLRLHLLVEDAGPTPLVSGLLAASGLLFTLVLVTPRAQRLREGEPGVHPRGV 158  
QY 158 EQRNAGLAWLR-----ORHQHQAQPGVLFADDDNTYSLELFOEMRTTKVSVWPV 209  
DB 159 EQRNKALDWLRGGVAGGKDPPTGQTGVVTFADDDNTYSLELFEEMRWTKGVSWPV 218  
QY 210 GLVGGRRYERPLVNGKVGWYTGWRADRPFAIDMAGFVLSQVILSNPKAVFKRRGSQP 269  
DB 219 GLVGLGRFEPQVQDGRVVGFTAWEPSPRFPFVDMAGFVALPDLKPNQAFDSTAPR- 277  
QY 270 GMESEFLKQITVVELEPKANNCTKVLVWHTTEKUNLANE 311  
DB 278 GHLESSLLSLVDPKPLEPAAANCTRVLVWHTTRTERPKMKQE 319  
RESULT 13  
ADY70386

ID ADY70386 standard; protein; 335 AA.  
XX ADY70386;  
XX 02-JUN-2005 (first entry)  
XX Galactosylgalactosylxylosylprotein 3-beta-glucuronosyltransferase 3.  
XX beta-amyloid; amyloid precursor protein; neuroprotective; nootropic;  
KW neurodegenerative disease; Alzheimers disease.  
XX Homo sapiens.  
XX WO2005023858-A1.  
XX 17-MAR-2005.  
XX 10-DEC-2003; 2003WO-EP013980.  
XX 05-SEP-2003; 2003EP-00019642.  
XX (CELL-) CELLZONE AG.  
XX Bouwmeester T, Drewes G, Hopf C, Joberty G, Rowley A;  
XX WPI; 2005-242127/25.  
XX Protein complex of beta-amyloid precursor protein processing pathway,  
PT useful as target for active agent of pharmaceutical, in treating or  
PT preventing Alzheimer's disease and related neurodegenerative disorders.  
XX Claim 1; SEQ ID NO 118; 1294pp; English.  
XX The invention relates to novel protein complexes of the beta-amyloid  
CC precursor protein processing pathway. The invention further comprises:  
CC component proteins of said complexes, fragments and derivatives of the  
CC component proteins and antibodies specific to the complexes. The  
CC invention also relates to methods for the use of the complexes of the APP  
CC processing pathway and their interacting proteins in screening, diagnosis  
CC and therapy, as well as to methods for preparing the complexes. The novel  
CC protein complexes have neuroprotective and nootropic activities. The  
CC protein complex, protein chosen from the proteins of Table 1 and  
CC antibodies are useful for diagnosing diseases and disorders,  
CC preferentially for diseases or disorders such as neurodegenerative  
CC disease such as Alzheimer's disease and related neurodegenerative  
CC disorders. The method is useful for treating or preventing a disease or  
CC disorder characterized by an aberrant amount of, activity of, component  
CC composition of or intracellular localization of the protein complex. The  
CC disease or disorder involves decreased levels of the amount or activity  
CC of the complex, or increased levels of the amount or activity of the  
CC protein complex. The protein complex and/or proteins listed in table 1,  
CC fifth column of the complex, is useful as a target for an active agent of  
CC a pharmaceutical, preferably a drug target in the treatment or prevention  
CC of a disease or disorder, preferentially of a disease or disorder such as  
CC neurodegenerative disease such as Alzheimer's disease and related  
CC neurodegenerative disorders. The pharmaceutical composition and method  
CC are useful for treating diseases and disorders, preferentially for  
CC diseases or disorders such as neurodegenerative disease such as  
CC Alzheimer's disease and related neurodegenerative disorders. This  
CC sequence represents one of the proteins used in a complex of the beta-  
CC amyloid precursor protein processing pathway of the invention.  
XX SQ Sequence 335 AA;  
Query Match 37.8%; Score 645; DB 9; Length 335;  
Best Local Similarity 43.6%; Pred. No. 3.3e-55;  
Matches 149; Conservative 41; Mismatches 94; Indels 58; Gaps 10;  
QY 1 MKSALFTRFFILLPWILVIMLDVDTTRP---VPPLTPRPYSPYAVGRGGA-----50  
DB 5 LKNVFLAYFLVSIAGLLVALVQLG---QPCDCLPPL-----RAAAEQLRQKD 48  
QY 51 -----RLPLRRGGPAHGTQKRNQSRPQPQPEQ-LPTIYAITPTYSRPVQKAELTRLAN 103

Db	49	LRISQOAEALRRPPPA-----PAQPEPEALPTIYVVTYARLVQKAEVLRLSQ	98
Qy	104	TFROVAOLHWLTVEDAAARSELVSRLARAGLPSTHLHVTPR-----RYKRPGL--PRAT	157
Db	99	TLSVLPRHLMLVEDAEGPTPLVSGLLAASGLLTHLVLTTPKAQRLUREGEPGWVHRPGV	158
Qy	158	EQRNAGLAWLR-----QRHQHQAQPGVLPFADDDNTYSLELFOEMRTTRKVSVPV	209
Db	159	EQRNKALDWLRGRGAVGGEKPPPGTQGVVYFADDDNTYSRELSEMEWTRGVSWPV	218
Qy	210	GLVGRYERPLVNGKVGWYGTWRADRPFAIDMAGFAVSLQVILSNPKAVFKRRGSQP	269
Db	219	GLVGLRFEQPGVQDGRVGFHTAWEPSPRPFPVDMAGFAVALPLLLDKPNAQFDSTAPR-	277
Qy	270	GMQSSDFLKQITTVVELEPKANNCTKVLVWHTRTKVNLANE	311
Db	278	GHLESSLLSHLVDPKDLPEAANCTTRVLVWHTRTKPKMKQ	319
RESULT 14			
AD	AD100832	standard; protein; 306 AA.	
AC	AD100832;		
XX	XX		
DT	DT	22-APR-2004 (first entry)	
DE	DE	Fruit fly DmGlcAT-I transglucuronidase (glucuronyltransferase) protein.	
XX	XX		
KW	KW	transglucuronidase; glucuronyltransferase; GAG; glucosaminoglycan chain;	
KW	KW	HNK-1 epitope; glycotecnology; fruit fly; DmGlcAT-I; enzyme.	
XX	XX		
OS	OS	Drosophila melanogaster.	
XX	XX		
PN	PN	WO2004003206-A1.	
XX	XX		
PD	PD	08-JAN-2004.	
XX	XX		
PF	PF	30-JUN-2003; 2003WO-JP008256.	
XX	XX		
PR	PR	01-JUL-2002; 2002JP-00192467.	
XX	XX		
PA	PA	(NISC-) JAPAN SCI & TECHNOLOGY CORP.	
PI	PI	Sugahara K, Kitagawa H;	
XX	XX		
DR	DR	WPI; 2004-083061/08.	
DR	DR	N-PSDB; AD100831.	
XX	XX		
PT	PT	Drosophila melanogaster-originated transglucuronidases and their encoded	
PT	PT	genes active in syntheses of GAG chain and HNK-1 epitope, useful as	
PT	PT	reagent in their studies and in glycotecnology.	
XX	XX		
PS	PS	Claim 1; SEQ ID NO 2; 75pp; Japanese.	
XX	XX		
CC	CC	The invention relates to a novel gene encoding a transglucuronidase	
CC	CC	(glucuronyltransferase) protein. The enzymes of the invention, DmGlcAT-1,	
CC	CC	DmGlcAT-PI and DmGlcAT-PII, and their encoded genes participate in the	
CC	CC	synthesis of the GAG (glucosaminoglycan) chain and HNK-1 epitope, both of	
CC	CC	which are particularly useful as reagents and in glycotecnology. The	
CC	CC	current sequence is that of the fruit fly DmGlcAT-I transglucuronidase	
CC	CC	(glucuronyltransferase) protein of the invention.	
XX	XX		
SQ	Sequence 306 AA;		
Query Match			
Best Local Similarity 35.7%; Score 608; DB 8; Length 306;			
Matches 136; Conservative 28; Mismatches 74; Indels 34; Gaps 9;			
Qy	66	RNOSRPOQPE-----POLPTIYAITPTYSRPVQKAELTRLANTRQVLAQLHWILV	116
Db	27	RNGKRTCGFEYLQAMFVGQDGLTPTIYALTPTYPRAQKAELTRLSHLFLLPHLHWIIV	86
Query Match			
Best Local Similarity 50.08; Pred. No. 1.4e-51;			
Matches 136; Conservative 28; Mismatches 74; Indels 34; Gaps 9;			
Qy	117	EDAAARSELVSRLARAGLP--STHLHVTPRPRYKRPG-----LPRATEQRNAGLAWLR	168
Db	87	EDNTATTPLVNLLDRAGLEKRSSTLNIKTPSEFLKGDPNWIKRPGVEQRNALLAWLR	146
Qy	169	-----QRHQHQAQPGVLPFADDDNTYSLELFOEMRTTR--KVSVPVGLVGGRRVERPL	221
Db	147	NHVDVDRH-----SIVFFMDDDNSYSTELFAEMSKIERGRVGVMPVGLVGGLMVERPL	199
Qy	222	V-ENG-KVVGWYGTWRADRPFAIDMAGFAVSLQVILSNPKAVFKRRGSQPMOESDFLKQ	279
Db	200	LTEDGKTKVGFENAMRPERPFPIDMAFAISMDLFI RNQATFSYE-VQGYOESEILRH	258
Qy	280	ITTVVELEPKANNCTKVLVWHTRTKVNLANE	311
Db	259	LTTRDQLQPLANRCTDVLVWHTRTKTKLAAE	290
RESULT 15			
AD	AD93879	standard; protein; 241 AA.	
XX	XX		
AC	AD93879;		
XX	XX		
DT	DT	06-NOV-2001 (first entry)	
XX	XX		
DE	DE	Human polypeptide, SEQ ID NO: 3997.	
XX	XX		
KW	KW	Human; full length cDNA; cDNA synthesis; oligo-capping.	
XX	XX		
OS	OS	Homo sapiens.	
XX	XX		
PN	PN	EF1130094-A2.	
XX	XX		
PD	PD	05-SEP-2001.	
XX	XX		
PF	PF	07-JUL-2000; 2000EP-00114089.	
XX	XX		
PR	PR	08-JUL-1999; 99JP-00194486.	
PR	PR	11-JAN-2000; 2000JP-00118774.	
XX	XX		
PR	PR	02-MAY-2000; 2000JP-00183765.	
XX	XX		
PA	PA	(HELI-) HELIX RES INST.	
XX	XX		
PI	PI	Ota T, Nishikawa T, Isogai T, Hayaishi K, Ishii S, Kawai Y;	
PI	PI	Wakamatsu A, Sugiyama T, Nagai K, Kojima S, Otsuki T, Koga H;	
XX	XX		
DR	DR	WPI; 2001-524255/58.	
DR	DR	N-PSDB; AAK94839.	
XX	XX		
PT	PT	830 Primers useful for synthesizing full length cDNA clones and their use	
PT	PT	in genetic manipulation.	
XX	XX		
PS	PS	Claim 8; SEQ ID NO 3997; 1380pp + Sequence Listing; English.	
XX	XX		
CC	CC	The invention relates to primers for synthesising full length cDNA	
CC	CC	clones. 830 cDNA molecules encoding a human protein have been isolated	
CC	CC	and nucleotide sequences of 5'- and 3'-ends of the cDNA molecules have	
CC	CC	been determined. Primers for synthesising the full length cDNA are useful	
CC	CC	for clarifying the function of the protein encoded by the cDNA. The full	
CC	CC	length clones were obtained by construction of full length enriched cDNA	
CC	CC	libraries that were synthesised by the oligo-capping method. The primers	
CC	CC	enable the production of the full length cDNA easily without any special	
CC	CC	methods. The present sequence is a polypeptide encoded by a full length	
CC	CC	human cDNA of the invention. Note: The sequence data for this patent did	
CC	CC	not form part of the printed specification, but was obtained in CD-ROM	
CC	CC	format directly from EPO	
XX	XX		
SQ	Sequence 241 AA;		
Query Match			
Best Local Similarity 29.9%; Score 510; DB 4; Length 241;			
Matches 114; Conservative 28; Mismatches 74; Indels 26; Gaps 5;			

Qy	77	PQLPTIYAITPTYSRPVQKAELTRLANTFRQVAQLHWILVEDAAARSELVSRFLARAGLP	136
Db	5	PPMGGWY-----KAEVLRLSQTLSLVRHLHLLVEDAEGPTPLVSGLLAASGLL	53
Qy	137	STHLHVTPTR---RYKRPGI--PRATEORNAGLAWLR-----QRHQHORAQPGVLF	182
Db	54	FTHLVVLTPKAQRLREGEFGWHPRGVEQRNKALDWLRGRGGAVGGEKDPPTPGTQGVVY	113
Qy	183	FADDDNTYSLELFQEMRTTRKVSVPVGLVGGRRYERPLVENGKVYVGYTGWRA DRPPAI	242
Db	114	FADDDNTYSRELFEEMRWTRGVSVMPVGLVGLRFEFGPQVQDGRVVGFHTAWEPSRPPV	173
Qy	243	DMAGFAVSLQVILSNPKAVFKRRGSPQMGQESDFLKQITTVBELEPKANNCTKVLVWHTR	302
Db	174	DMAGFAVALPFLLDKPNQAFDSTAPR-GHLESLSLHLVDPKDLEPRAANCTRSLAVSPR	232
Qy	303	TE	304
Db	233	LE	234

Search completed: May 9, 2006, 10:53:26  
Job time : 90 secs



GenCore version 5.1.8  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: May 9, 2006, 10:54:58 ; Search time 47 Seconds  
(without alignments)  
568.175 Million cell updates/sec

Title: US-10-634-905-2  
Perfect score: 1705  
Sequence: 1 MKSALTRFFILLPWILVI.....EKNVLANEPKYHLDTVKIEV 323

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 572060 seqs, 82675679 residues  
Total number of hits satisfying chosen parameters: 572060

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents AA:\*  
1: /cgn2\_6/ptodata/1/iaa/5\_COMB.pep.\*  
2: /cgn2\_6/ptodata/1/iaa/6\_COMB.pep.\*  
3: /cgn2\_6/ptodata/1/iaa/H\_COMB.pep.\*  
4: /cgn2\_6/ptodata/1/iaa/PTUS\_COMB.pep.\*  
5: /cgn2\_6/ptodata/1/iaa/RE\_COMB.pep.\*  
6: /cgn2\_6/ptodata/1/iaa/backfiles.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1705	100.0	323	2	US-09-816-095-2
2	1526.5	89.5	324	2	US-09-816-095-4
3	1039	60.9	197	2	US-09-816-095-5
4	777.5	45.6	347	2	US-09-059-369-2
5	486.5	28.5	291	2	US-09-270-767-33192
6	486.5	28.5	291	2	US-09-270-767-48409
7	236	13.8	142	2	US-09-270-767-45263
8	130.5	7.7	521	2	US-09-252-991A-18119
9	115.5	6.8	603	2	US-09-252-991A-25264
10	111	6.5	666	2	US-09-050-739-70
11	107	6.3	396	2	US-09-252-991A-26907
12	107	6.3	1037	2	US-09-252-991A-25361
13	106.5	6.2	475	2	US-09-252-991A-25837
14	106.5	6.2	720	2	US-09-252-991A-23212
15	105.5	6.2	388	2	US-09-252-991A-17917
16	104	6.1	470	2	US-09-252-991A-19467
17	104	6.1	618	2	US-09-252-991A-23696
18	103	6.0	25	2	US-09-059-369-4
19	102	6.0	1418	2	US-09-252-991A-33367
20	99.5	5.8	436	2	US-09-252-991A-20220
21	99.5	5.8	506	2	US-09-252-991A-18165
22	99.5	5.8	560	2	US-09-252-991A-26107
23	99	5.8	361	2	US-09-252-991A-24065
24	98	5.7	288	2	US-09-252-991A-27950
25	97.5	5.7	1706	2	US-09-252-991A-31760
26	97	5.7	485	2	US-09-252-991A-17141
27	97	5.7	618	2	US-09-252-991A-28358

28	96.5	5.7	481	2	US-09-252-991A-23317	Sequence 23317, A
29	96.5	5.7	628	2	US-09-252-991A-24741	Sequence 24741, A
30	96	5.6	159	2	US-09-252-991A-26509	Sequence 26509, A
31	95.5	5.6	194	2	US-09-252-991A-22313	Sequence 22313, A
32	95.5	5.6	312	2	US-09-252-991A-20216	Sequence 20216, A
33	95.5	5.6	440	2	US-09-252-991A-24173	Sequence 24173, A
34	95.5	5.6	594	2	US-09-252-991A-20345	Sequence 20345, A
35	95.5	5.6	871	2	US-09-252-991A-19431	Sequence 19431, A
36	95	5.6	661	2	US-09-252-991A-18960	Sequence 18960, A
37	94.5	5.5	381	2	US-09-252-991A-23050	Sequence 23050, A
38	94.5	5.5	437	2	US-09-252-991A-26942	Sequence 26942, A
39	94.5	5.5	458	2	US-09-252-991A-28585	Sequence 28585, A
40	94.5	5.5	582	2	US-09-252-991A-25366	Sequence 25366, A
41	94.5	5.5	780	2	US-09-949-016-10205	Sequence 10205, A
42	94.5	5.5	875	2	US-09-252-991A-30056	Sequence 30056, A
43	94.5	5.5	1245	2	US-09-252-991A-30935	Sequence 30935, A
44	93.5	5.5	160	2	US-09-621-976-4217	Sequence 4217, Ap
45	93.5	5.5	254	2	US-09-248-796A-15686	Sequence 15686, A

ALIGNMENTS

RESULT 1  
US-09-816-095-2  
; Sequence 2, Application US/09816095  
; Patent No. 6664084  
; GENERAL INFORMATION:  
; APPLICANT: GAN, Weiniu  
; TITLE OF INVENTION: ISOLATED HUMAN ENZYME PROTEINS, NUCLEIC  
; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN ENZYME PROTEINS, AND USES  
; FILE REFERENCE: CLO01147  
; CURRENT APPLICATION NUMBER: US/09/816, 095  
; CURRENT FILING DATE: 2001-03-26  
; NUMBER OF SEQ ID NOS: 5  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 2  
; LENGTH: 323  
; TYPE: PRT  
; ORGANISM: Human  
US-09-816-095-2

Query Match		100.0%;	Score 1705;	DB 2;	Length 323;
Best Local Similarity		100.0%;	Pred. No. 7e-174;		
Matches 323;		Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	1	MKSALTRFFILLPWILVIIMLDVDTTRRPVPLTPRPYFSPYAVGCGGARLPLRRGGPA	60		
DB	1	MKSALTRFFILLPWILVIIMLDVDTTRRPVPLTPRPYFSPYAVGCGGARLPLRRGGPA	60		
QY	61	HGTOKENQSRPOPEPOLPTIYAITPTYSRPVQKAELTRLANTFRVQALHWILVEDAA	120		
DB	61	HGTOKENQSRPOPEPOLPTIYAITPTYSRPVQKAELTRLANTFRVQALHWILVEDAA	120		
QY	121	ARSELVSRFLARAGLPSTHLVPTPRYKRPGLPRATEQORNAGLWLRQHQHQRQPGV	180		
DB	121	ARSELVSRFLARAGLPSTHLVPTPRYKRPGLPRATEQORNAGLWLRQHQHQRQPGV	180		
QY	181	LPFADDNTYSLELFOEMRTTRKVSVPVGLVGGRRYERPLVENGKVGYTGWADRPFF	240		
DB	181	LPFADDNTYSLELFOEMRTTRKVSVPVGLVGGRRYERPLVENGKVGYTGWADRPFF	240		
QY	241	AIDMAGFVSLQVILSNPKAVFKRGSQPMQESDFLKQITTVVEELEPKANNCTKVLVWH	300		
DB	241	AIDMAGFVSLQVILSNPKAVFKRGSQPMQESDFLKQITTVVEELEPKANNCTKVLVWH	300		
QY	301	TRTEKVNLANEPKYHLDTVKIEV	323		
DB	301	TRTEKVNLANEPKYHLDTVKIEV	323		

RESULT 2

```
US-09-816-095-4
; Sequence 4, Application US/09816095
; Patent No. 6664084
; GENERAL INFORMATION:
; APPLICANT: GAN, Weiniu
; TITLE OF INVENTION: ISOLATED HUMAN ENZYME PROTEINS, NUCLEIC
; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN ENZYME PROTEINS, AND USES
; TITLE OF INVENTION: THEREOF
; FILE REFERENCE: CL001147
; CURRENT APPLICATION NUMBER: US/09/816, 095
; CURRENT FILING DATE: 2001-03-26
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 324
; TYPE: PRT
; ORGANISM: Rattus norvegicus
US-09-816-095-4

Query Match      89.5%; Score 1526.5; DB 2; Length 324;
Best Local Similarity 89.8%; Pred. No. 8.9e-155;
Matches 291; Conservative 10; Mismatches 22; Indels 1; Gaps 1;

QY 1 MKSALTRFFILLPWILVIMLDVTRRPVPLTPRPYFSPYAVGRGGARLPLRRGGPA 60
DB 1 MKSALCNRFILLPWILVIMLDVDRPPAPQLTSRPYFSPHTVCGGSRVPLRRSSPG 60
QY 61 H-GTQKRNQSRPQPEPQLPTIYAITPTYSRVPQKAELTRANTFRQVAQLHWILVEDA 119
DB 61 RDAAEKNESRPLQPEPLPTIYAITPTYSRVPQKAELTRANTFRQVAQLHWILVEDR 120
QY 120 AARSELSRFLARAGLPSTHLHVPTPRRYKRPGLPRATEORNAGLAWLRQHORHQAQPG 179
DB 121 ATRSELVSSFLARAGLPSTHLHVPTPRRYKRPGLPRATEORNAGLAWLRQHORHQAQPG 180
QY 180 VLFADDNTYSLELFOEMRTTKVSWPVLGVRGRYERPLVNGKVGWYTGWRADRP 239
DB 181 VLFADDNTYSLELFOEMRTTKVSWPVLGVRGRYERPLVNGKVGWYTGWRADRP 240
QY 240 FAIDMAGFAVSLQVILSNKAVFKRRGSGQFMQESDFLKQITTVLELEPKANCTKVLVW 299
DB 241 FAIDMAGFAVSLQVILSNKAVFKRRGSGQFMQESDFLKQITTVLELEPKANCTKVLVW 300
QY 300 HTRTEKNLANEPKYHLDTVKIEV 323
DB 301 HTRTEKNLANEPKYHMDTVNIEV 324

RESULT 3
US-09-816-095-5
; Sequence 5, Application US/09816095
; Patent No. 6664084
; GENERAL INFORMATION:
; APPLICANT: GAN, Weiniu
; TITLE OF INVENTION: ISOLATED HUMAN ENZYME PROTEINS, NUCLEIC
; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN ENZYME PROTEINS, AND USES
; TITLE OF INVENTION: THEREOF
; FILE REFERENCE: CL001147
; CURRENT APPLICATION NUMBER: US/09/816, 095
; CURRENT FILING DATE: 2001-03-26
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5
; LENGTH: 197
; TYPE: PRT
; ORGANISM: Human
US-09-816-095-5

Query Match      60.9%; Score 1039; DB 2; Length 197;
Best Local Similarity 100.0%; Pred. No. 6.3e-103;
Matches 197; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKSALTRFFILLPWILVIMLDVTRRPVPLTPRPYFSPYAVGRGGARLPLRRGGPA 60
DB 1 MKSALTRFFILLPWILVIMLDVTRRPVPLTPRPYFSPYAVGRGGARLPLRRGGPA 60
QY 61 HGTQKRNQSRPQPEPQLPTIYAITPTYSRVPQKAELTRANTFRQVAQLHWILVEDAA 120
DB 61 HGTQKRNQSRPQPEPQLPTIYAITPTYSRVPQKAELTRANTFRQVAQLHWILVEDAA 120
QY 121 ARSELVSRFLARAGLPSTHLHVPTPRRYKRPGLPRATEORNAGLAWLRQHORHQAQPGV 180
DB 121 ARSELVSRFLARAGLPSTHLHVPTPRRYKRPGLPRATEORNAGLAWLRQHORHQAQPGV 180
QY 181 LFFADDNTYSLELFOE 197
DB 181 LFFADDNTYSLELFOE 197

US-09-059-369-2
; Sequence 2, Application US/09059369
; Patent No. 6040156
; GENERAL INFORMATION:
; APPLICANT: KAWASAKI, TOSHISUKE
; APPLICANT: OKA, SHOGO
; TITLE OF INVENTION: DNA ENCODING GLUCURONYLTRANSFERASE
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSES: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,
; ADDRESSES: P.C.
; STREET: 1755 S. JEFFERSON DAVIS HIGHWAY, FOURTH FLOOR
; CITY: ARLINGTON
; STATE: VA
; COUNTRY: USA
; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/059,369
; FILING DATE: 14-APR-1998
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 9-127065
; FILING DATE: 16-MAY-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: OBLON, NORMAN F.
; REGISTRATION NUMBER: 24,618
; REFERENCE/DOCKET NUMBER: 9378-0002-0
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703-413-3000
; TELEFAX: 703-413-2220
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 347 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-059-369-2

Query Match      45.6%; Score 777.5; DB 2; Length 347;
Best Local Similarity 48.3%; Pred. No. 1.4e-74;
Matches 167; Conservative 36; Mismatches 86; Indels 57; Gaps 8;

QY 11 ILLPWILVII-----MLDV-----DTRRPVPP-LTRPYF-----SPYA 44
DB 26 IVLPWTILITVWHQSSLAFLAVHDKGSDPRHEAPGADREYCMDSRDIYEVVTV 85
QY 45 VGRGGARLPLRRGGPAHGTQKRNQSRPQPEPQLPTIYAITPTYSRVPQKAELTRANT 104
DB 86 -----YTRPPFWS-D-TLPTIHVVTPYTSRVPQKAELTRMANT 121
QY 105 FRQVAQLHWILVEDAAARSELVSRFLARAGLPSTHLHVPTPRYK-----RPLGRATE 158
```

Db 122 LLHVPNLHVLVEDAPRRPTPLTABLLRDTGLNTHLVETPRNKLRGDARDPRIPRGTM 181  
Qy 159 ORNAGLAWLRQHQRHQAQPCVLPFADDDNTYSLELFOEMRTTRKVSVPVGLVGGRYE 218  
Db 182 ORNLALRWLRETFPRNSTQGVVVFADDDNTYSLELFEEMKSTRVSVPVAFVGGURYE 241  
Qy 219 RPLVEN-GKVVGVWTGWDRAPFAIDMAGFAVSLQVILSNPKAVFKRGSGQMGQESDFL 277  
Db 242 APRVNGAGKVVGMKTVDPHPFPALDMAGFAVNLRLILQRSQAYFKLRGVKGQYQESSLL 301  
Qy 278 KQITVVELEPKANNCTKVLVWHTRTKVKNLANEPKXHLDTVKIEV 323  
Db 302 RELVTLNDLEPKAANCTKILVWHTRTKPKVLVNEGKGFDTDPSEI 347

## RESULT 5

US-09-270-767-33192  
; Sequence 33192, Application US/09270767  
; Patent No. 6703491  
; GENERAL INFORMATION:  
; APPLICANT: Homburger et al.  
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster  
; FILE REFERENCE: File Reference: 7326-094  
; CURRENT APPLICATION NUMBER: US/09/270,767  
; CURRENT FILING DATE: 1999-03-17  
; NUMBER OF SEQ ID NOS: 62517  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 33192  
; LENGTH: 291  
; TYPE: PRT  
; ORGANISM: Drosophila melanogaster  
US-09-270-767-33192

Query Match 28.5%; Score 486.5; DB 2; Length 291;  
Best Local Similarity 40.7%; Pred. No. 1.5e-43;  
Matches 103; Conservative 45; Mismatches 90; Indels 15; Gaps 7;  
Qy 68 QSRPQPPEPOLPIYAITPTYSRPVQKAEITRLANTFRQVAQLHWTILVEDAAARSELVS 127  
Db 2 QDKPQ-SDYQVLPVIYFTPTYPREQIPELTRLAHTLLHIPRLHVLVADDEKCNNDYMD 60  
Qy 128 RFLARAGLPSTHLHVPTRRYK--RPGLPRAEQNAGLAWLRQHQRHQAQPCVLPFAD 185  
Db 61 TLLYRFGMPFTHMVSMPKSRNEKPA-PRGVANRAALQWIRQHN---LTNGILYFGD 115  
Qy 186 DDNTYSLELFOEMRTTRKVSVPVGLVGGRRYERPLVNGKVGWYTGWRADRPFAIDMA 245  
Db 116 DDNTYDLRLFEIRKQTVSMFPVGLIADYGVSGPVVRKGVAFVFLDSWVAGRRWPVDMA 175  
Qy 246 GFAVSLQVILSNPKAVFKRGSGQMGQESDPLKQI-TTVELEPKANNCTKVLVWHTRTE 304  
Db 176 GFAVNLEYMAQYP---YVNMPYKPGYEDDLFLRSIGLQMLNLEBPRGNCTEILVWHTQTK 232  
Qy 305 KVNLAN---EPKY 314  
Db 233 SKKLGWVRLESKY 245

## RESULT 6

US-09-270-767-48409  
; Sequence 48409, Application US/09270767  
; Patent No. 6703491  
; GENERAL INFORMATION:  
; APPLICANT: Homburger et al.  
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster  
; FILE REFERENCE: File Reference: 7326-094  
; CURRENT APPLICATION NUMBER: US/09/270,767  
; CURRENT FILING DATE: 1999-03-17  
; NUMBER OF SEQ ID NOS: 62517  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 48409  
; LENGTH: 291

; TYPE: PRT  
; ORGANISM: Drosophila melanogaster  
US-09-270-767-48409  
Query Match 28.5%; Score 486.5; DB 2; Length 291;  
Best Local Similarity 40.7%; Pred. No. 1.5e-43;  
Matches 103; Conservative 45; Mismatches 90; Indels 15; Gaps 7;  
Qy 68 QSRPQPPEPOLPIYAITPTYSRPVQKAEITRLANTFRQVAQLHWTILVEDAAARSELVS 127  
Db 2 QDKPQ-SDYQVLPVIYFTPTYPREQIPELTRLAHTLLHIPRLHVLVADDEKCNNDYMD 60  
Qy 128 RFLARAGLPSTHLHVPTRRYK--RPGLPRAEQNAGLAWLRQHQRHQAQPCVLPFAD 185  
Db 61 TLLYRFGMPFTHMVSMPKSRNEKPA-PRGVANRAALQWIRQHN---LTNGILYFGD 115  
Qy 186 DDNTYSLELFOEMRTTRKVSVPVGLVGGRRYERPLVNGKVGWYTGWRADRPFAIDMA 245  
Db 116 DDNTYDLRLFEIRKQTVSMFPVGLIADYGVSGPVVRKGVAFVFLDSWVAGRRWPVDMA 175  
Qy 246 GFAVSLQVILSNPKAVFKRGSGQMGQESDPLKQI-TTVELEPKANNCTKVLVWHTRTE 304  
Db 176 GFAVNLEYMAQYP---YVNMPYKPGYEDDLFLRSIGLQMLNLEBPRGNCTEILVWHTQTK 232  
Qy 305 KVNLAN---EPKY 314  
Db 233 SKKLGWVRLESKY 245

## RESULT 7

US-09-270-767-45263  
; Sequence 45263, Application US/09270767  
; Patent No. 6703491  
; GENERAL INFORMATION:  
; APPLICANT: Homburger et al.  
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster  
; FILE REFERENCE: File Reference: 7326-094  
; CURRENT APPLICATION NUMBER: US/09/270,767  
; CURRENT FILING DATE: 1999-03-17  
; NUMBER OF SEQ ID NOS: 62517  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 45263  
; LENGTH: 142  
; TYPE: PRT  
; ORGANISM: Drosophila melanogaster  
US-09-270-767-45263

Query Match 13.8%; Score 236; DB 2; Length 142;  
Best Local Similarity 35.3%; Pred. No. 3.4e-17;  
Matches 48; Conservative 29; Mismatches 45; Indels 14; Gaps 4;  
Qy 197 EMRTTRKVSVPVGLVGGRRYERPLVNGKVGWYTGWRADRPFAIDMAGFAVSLQVILS 256  
Db 8 QMRYISKVAMVPVGLVTKTVSGVSPITQAGKLVGYDGVIGRKYVPVDMAGFAVSVKFLKE 67  
Qy 257 NPKAVFKRGSGQMGQESDPLKQITTVE--ELEPKANNCTKVLVWHTRTEK---VNLANE 311  
Db 68 RPNA---QMPFKPGYEDGFLRSIAPLDADAEIELLADECRDILTWTQTQKPAQAQALNR 124  
Qy 312 PKY-----HLDTVKI 321  
Db 125 TRYKNTNLEHIDRLLV 140

## RESULT 8

US-09-252-991A-18119  
; Sequence 18119, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 107196.136





```
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 23212
; LENGTH: 720
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-23212

Query Match          6.2%; Score 106.5; DB 2; Length 720;
Best Local Similarity 27.7%; Pred.No.0.029;
Matches 53; Conservative 16; Mismatches 69; Indels 53; Gaps 9;

Qy 29 RPVPPLTP--RPYFSPYA-----VGRGARLPLRRGGPAHGCTQRNOSRRPOQPEPQL 79
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 121 RPAPRLPALRPY--PGARGARRTAGRAG-RQPLAGAMGAGAQAHR-----ANFGPAL 171

Qy 80 PT-----IYAITPTYSRP-----VQKAELTRLANTFQVAQLHLV-----116
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 172 PARTGHDLRMAPTPALPRGHRGAGSRQGDPAADRLRPLQQPLGVHLDPVPPAGRHHTGTI 231

Qy 117 -----EDAAARGSELVSRFLARAGLPSTLHLVPTPRYKRPGLPRATEQRNAGLAWL 167
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 232 PPGQPRPGHADAVTRACVAAMLAAATIRSSHANRQPPSEHPRGVPRLPQ---AGLQRL 288

Qy 168 RQRHQHQAQP 178
   | : | : |
Db 289 RRAHPPRLFP 299

RESULT 15
US-09-252-991A-17917
; Sequence 17917, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 17917
; LENGTH: 388
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-17917

Query Match          6.2%; Score 105.5; DB 2; Length 388;
Best Local Similarity 28.0%; Pred.No.0.014;
Matches 45; Conservative 16; Mismatches 61; Indels 39; Gaps 8;

Qy 20 IIMLDVTRRPVPPPLTPRPYPFYVGRGARLPLRRGGPAHGCTQRNOSRRPOQPEPQL 79
   :| : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 80 LVNIRLHEHTPDPAPAPATWRA-AVRRCGPRIQRRTG-SHGRQQAQTRLPGRPP--135

Qy 80 PTIYAITPTYSRVPQKAELTRLANTFQVAQLHLVDEDAARSELVSR-FLARAGLPST 138
   :| : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 136 -----AAAAGAARPDRR-----VFRQSSQAP--AADQPVARARLFQRHQGTAGSLGT 181

Qy 139 HLHVPTPRYKRPGLPRATEQRNAGLAWLQRHQHQAQP 179
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 182 G---PRPRPCARPG-----PGDHRRVBFG 202
```

GenCore version 5.1.8  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: May 9, 2006, 10:55:28 ; Search time 165 Seconds  
(without alignments)  
817.933 Million cell updates/sec

Title: US-10-634-905-2

Perfect score: 1705

Sequence: 1 MKSALFTRFFILLPWILVI.....EKNVLANEPKYHLDTVKIEV 323

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1867569 seqs, 417829326 residues

Total number of hits satisfying chosen parameters: 1867569

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- Published Applications AA\_Main:\*  
1: /cgn2\_6/prodata/1/pubpaa/US07\_PUBCOMB.pgp:\*  
2: /cgn2\_6/prodata/1/pubpaa/US08\_PUBCOMB.pgp:\*  
3: /cgn2\_6/prodata/1/pubpaa/US09\_PUBCOMB.pgp:\*  
4: /cgn2\_6/prodata/1/pubpaa/US10\_PUBCOMB.pgp:\*  
5: /cgn2\_6/prodata/1/pubpaa/US10B\_PUBCOMB.pgp:\*  
6: /cgn2\_6/prodata/1/pubpaa/US11\_PUBCOMB.pgp:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1705	100.0	323	3	US-09-816-095-2
2	1705	100.0	323	4	US-10-634-905-2
3	1526.5	89.5	324	3	US-09-816-095-4
4	1526.5	89.5	324	4	US-10-634-905-4
5	1039	60.9	197	3	US-09-816-095-5
6	1039	60.9	197	4	US-10-634-905-5
7	653	38.3	335	5	US-10-698-190-16
8	653	38.3	335	5	US-10-723-860-4238
9	502.5	29.5	479	6	US-11-097-143-13308
10	487.5	28.6	443	6	US-11-097-143-7506
11	422	24.8	145	4	US-10-080-254-93
12	422	24.8	145	4	US-10-242-355-637
13	406	23.8	691	5	US-10-450-763-30691
14	324.5	19.3	201	4	US-10-403-571-34
15	312.5	18.3	351	4	US-10-437-963-174074
16	305.5	17.9	331	4	US-10-425-115-330485
17	293	17.2	414	4	US-10-424-599-147426
18	256.5	15.0	446	4	US-10-437-963-135854
19	249	14.6	448	4	US-10-425-115-366742
20	246	14.4	451	4	US-10-425-115-292707
21	246	14.4	479	4	US-10-425-114-68877
22	244.5	14.3	528	4	US-10-425-115-360409
23	241.5	14.2	460	4	US-10-425-114-62647
24	229	13.4	524	4	US-10-437-963-177911
25	227	13.3	411	4	US-10-425-114-54143
26	221	13.0	385	4	US-10-425-115-345009
27	213	12.5	445	4	US-10-437-963-107957

28	210.5	12.3	342	4	US-10-424-599-178009	Sequence 178009,
29	209.5	12.3	316	4	US-10-425-114-53567	Sequence 53567, A
30	202.5	11.9	344	4	US-10-424-599-217815	Sequence 217815,
31	202.5	11.9	471	4	US-10-437-963-125866	Sequence 125866,
32	201.5	11.8	494	4	US-10-424-599-233474	Sequence 233474,
33	200	11.7	371	4	US-10-437-963-200415	Sequence 200415,
34	194.5	11.4	167	5	US-10-450-763-43912	Sequence 43912, A
35	188.5	11.1	367	4	US-10-425-115-290232	Sequence 290232,
36	188	11.0	549	4	US-10-437-963-102770	Sequence 102770,
37	181	10.6	235	4	US-10-437-963-140001	Sequence 140001,
38	179	10.5	883	4	US-10-425-115-296340	Sequence 296340,
39	171	10.0	273	4	US-10-425-114-50568	Sequence 50568, A
40	162	9.5	384	4	US-10-425-115-350198	Sequence 350198,
41	145	8.5	156	4	US-10-767-701-48703	Sequence 48703, A
42	141.5	8.3	378	4	US-10-425-114-42131	Sequence 42131, A
43	139	8.2	40	4	US-10-029-386-33145	Sequence 33145, A
44	138.5	8.1	374	4	US-10-425-115-364687	Sequence 364687,
45	137.5	8.1	267	4	US-10-425-114-54605	Sequence 54605, A

ALIGNMENTS

RESULT 1

US-09-816-095-2  
; Sequence 2, Application US/09816095  
; Patent No. US20020137164A1  
; GENERAL INFORMATION:  
; APPLICANT: GAN, Weiniu  
; TITLE OF INVENTION: ISOLATED HUMAN ENZYME PROTEINS, NUCLEIC  
; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN ENZYME PROTEINS, AND USES  
; FILE REFERENCE: CL001147  
; CURRENT APPLICATION NUMBER: US/09/816,095  
; CURRENT FILING DATE: 2001-03-26  
; NUMBER OF SEQ ID NOS: 5  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 2  
; LENGTH: 323  
; TYPE: PRT  
; ORGANISM: Human  
US-09-816-095-2

Query Match		100.0%;	Score 1705;	DB 3;	Length 323;
Best Local Similarity		100.0%;	Pred. No. 2.9e-149;		
Matches 323;		Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	1	MKSALFTRFFILLPWILVIIMLDVDRTRPVPLTPRPYFSPYAVGRGGARLPLRRGGPA	60		
DB	1	MKSALFTRFFILLPWILVIIMLDVDRTRPVPLTPRPYFSPYAVGRGGARLPLRRGGPA	60		
QY	61	HGTQKNSRQPOPEPQLPTIYAITPTYSRPVQKAELTRANTFRQVQLHWILVEDAA	120		
DB	61	HGTQKNSRQPOPEPQLPTIYAITPTYSRPVQKAELTRANTFRQVQLHWILVEDAA	120		
QY	121	ARSELVSRFLARAGLPSTHLHVTPTPRYKRPGLPRATEQRNAGLWLRQHRQHPQPGV	180		
DB	121	ARSELVSRFLARAGLPSTHLHVTPTPRYKRPGLPRATEQRNAGLWLRQHRQHPQPGV	180		
QY	181	LFADFDDNTYSLELFQEMRTTRKVSVPVGLVGGRRYERPLVGVKVGWYTCWRADRPF	240		
DB	181	LFADFDDNTYSLELFQEMRTTRKVSVPVGLVGGRRYERPLVGVKVGWYTCWRADRPF	240		
QY	241	AIDMAGFVSLQVILSNPKAVFKRGSGQMGSDFLKQITTTVEELEPKANNCTKVLVWH	300		
DB	241	AIDMAGFVSLQVILSNPKAVFKRGSGQMGSDFLKQITTTVEELEPKANNCTKVLVWH	300		
QY	301	TRTEKVLNANEPKYHLDTVKIEV	323		
DB	301	TRTEKVLNANEPKYHLDTVKIEV	323		

RESULT 2

```
US-10-634-905-2
; Sequence 2, Application US/10634905
; Publication No. US20040067225A1
; GENERAL INFORMATION:
; APPLICANT: GAN, Weiniu et al.
; TITLE OF INVENTION: ISOLATED HUMAN ENZYME PROTEINS, NUCLEIC
; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN ENZYME PROTEINS, AND USES
; TITLE OF INVENTION: THEREOF
; FILE REFERENCE: CL001147-DIV
; CURRENT APPLICATION NUMBER: US/10/634,905
; CURRENT FILING DATE: 2003-08-06
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 323
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-634-905-2

Query Match      100.0%; Score 1705; DB 4; Length 323;
Best Local Similarity 100.0%; Pred. No. 2.9e-149;
Matches 323; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKSALTRFFILLPWILVIMLDVDRPRPAQLTSRPFSPHTVCGGSRVPLRRSGP 60
DB 1 MKSALTRFFILLPWILVIMLDVDRPRPAQLTSRPFSPHTVCGGSRVPLRRSGP 60
QY 61 HGTQKNSRQPOPEPOLPTIYAITPTYSRPVQKAELTRANTFRQVAQLHWILVEDA 120
DB 61 HGTQKNSRQPOPEPOLPTIYAITPTYSRPVQKAELTRANTFRQVAQLHWILVEDA 120
QY 120 AARSELVSRFLARAGLPSTHLHVPTPRYKRPGLPRATEQORNAGLAWLRQRHQRAQPG 179
DB 120 AARSELVSRFLARAGLPSTHLHVPTPRYKRPGLPRATEQORNAGLAWLRQRHQRAQPG 179
QY 180 VLPFFADDDNTYSLELFOEMRTTRKVSVMVGLVGGRRYERPLVGVGWTGWRADRP 239
DB 180 VLPFFADDDNTYSLELFOEMRTTRKVSVMVGLVGGRRYERPLVGVGWTGWRADRP 239
QY 240 FAIDMAGFAVSLQVILSNPKAVFKRRGSGQPMQESDFLKQITTVDELEPKANNCTKVLVW 299
DB 240 FAIDMAGFAVSLQVILSNPKAVFKRRGSGQPMQESDFLKQITTVDELEPKANNCTKVLVW 299
QY 300 HTRTEKVNLANEPKHYLDTVKIEV 323
DB 300 HTRTEKVNLANEPKHYLDTVKIEV 323

RESULT 4
US-10-634-905-4
; Sequence 4, Application US/10634905
; Publication No. US20040067225A1
; GENERAL INFORMATION:
; APPLICANT: GAN, Weiniu et al.
; TITLE OF INVENTION: ISOLATED HUMAN ENZYME PROTEINS, NUCLEIC
; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN ENZYME PROTEINS, AND USES
; TITLE OF INVENTION: THEREOF
; FILE REFERENCE: CL001147-DIV
; CURRENT APPLICATION NUMBER: US/10/634,905
; CURRENT FILING DATE: 2003-08-06
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 324
; TYPE: PRT
; ORGANISM: Rattus norvegicus
US-10-634-905-4

Query Match      89.5%; Score 1526.5; DB 4; Length 324;
Best Local Similarity 89.8%; Pred. No. 1e-132;
Matches 291; Conservative 10; Mismatches 22; Indels 1; Gaps 1;

QY 1 MKSALTRFFILLPWILVIMLDVDRPRPAQLTSRPFSPHTVCGGSRVPLRRSGP 60
DB 1 MKSALTRFFILLPWILVIMLDVDRPRPAQLTSRPFSPHTVCGGSRVPLRRSGP 60
QY 61 HGTQKNSRQPOPEPOLPTIYAITPTYSRPVQKAELTRANTFRQVAQLHWILVEDA 119
DB 61 RDAAEKRNESRQLQPEPRPLPTIYAITPTYSRPVQKAELTRANTFRQVAQLHWILVEDR 120
QY 120 AARSELVSRFLARAGLPSTHLHVPTPRYKRPGLPRATEQORNAGLAWLRQRHQRAQPG 179
DB 120 AARSELVSRFLARAGLPSTHLHVPTPRYKRPGLPRATEQORNAGLAWLRQRHQRAQPG 179
QY 180 VLPFFADDDNTYSLELFOEMRTTRKVSVMVGLVGGRRYERPLVGVGWTGWRADRP 239
DB 180 VLPFFADDDNTYSLELFOEMRTTRKVSVMVGLVGGRRYERPLVGVGWTGWRADRP 239
QY 240 FAIDMAGFAVSLQVILSNPKAVFKRRGSGQPMQESDFLKQITTVDELEPKANNCTKVLVW 299
DB 240 FAIDMAGFAVSLQVILSNPKAVFKRRGSGQPMQESDFLKQITTVDELEPKANNCTKVLVW 300
QY 300 HTRTEKVNLANEPKHYLDTVKIEV 323
DB 300 HTRTEKVNLANEPKHYLDTVKIEV 323

US-09-816-095-2
; Sequence 4, Application US/09816095
; Patent No. US20020137164A1
; GENERAL INFORMATION:
; APPLICANT: GAN, Weiniu
; TITLE OF INVENTION: ISOLATED HUMAN ENZYME PROTEINS, NUCLEIC
; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN ENZYME PROTEINS, AND USES
; TITLE OF INVENTION: THEREOF
; FILE REFERENCE: CL001147
; CURRENT APPLICATION NUMBER: US/09/816,095
; CURRENT FILING DATE: 2001-03-26
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 324
; TYPE: PRT
; ORGANISM: Rattus norvegicus
US-09-816-095-2

Query Match      89.5%; Score 1526.5; DB 3; Length 324;
Best Local Similarity 89.8%; Pred. No. 1e-132;
Matches 291; Conservative 10; Mismatches 22; Indels 1; Gaps 1;

QY 1 MKSALTRFFILLPWILVIMLDVDRPRPAQLTSRPFSPHTVCGGSRVPLRRSGP 60
```



```
RESULT 5
US-09-816-095-5
; Sequence 5, Application US/09816095
; Patent No. US20020137164A1
; GENERAL INFORMATION:
; APPLICANT: GAN, Weinlu
; TITLE OF INVENTION: ISOLATED HUMAN ENZYME PROTEINS, NUCLEIC
; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN ENZYME PROTEINS, AND USES
; TITLE OF INVENTION: THEREOF
; FILE REFERENCE: CL001147
; CURRENT APPLICATION NUMBER: US/09/816,095
; CURRENT FILING DATE: 2001-03-26
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5
; LENGTH: 197
; TYPE: PRT
; ORGANISM: Human
US-09-816-095-5

Query Match          60.9%; Score 1039; DB 3; Length 197;
Best Local Similarity 100.0%; Pred. No. 7.7e-88;
Matches 197; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MKSALFTRFFILLPWILIVIMLDVTRRPVPLTPRPYFSPYAVGGRGRLPLRRGGPA 60
Db 1 MKSALFTRFFILLPWILIVIMLDVTRRPVPLTPRPYFSPYAVGGRGRLPLRRGGPA 60

Qy 61 HGTQKRNQSRPQPOPEQLPTIYAITYTSRPVQKAELTRANTFROVAQLHWILVEDAA 120
Db 61 HGTQKRNQSRPQPOPEQLPTIYAITYTSRPVQKAELTRANTFROVAQLHWILVEDAA 120

Qy 121 ARSELVSRFLARAGLPSTHLHVPTRRYKRPGLPRATEQRNAGLAWLRQRHQHORAQPGV 180
Db 121 ARSELVSRFLARAGLPSTHLHVPTRRYKRPGLPRATEQRNAGLAWLRQRHQHORAQPGV 180

Qy 181 LFFADDDNTYSLELFOE 197
Db 181 LFFADDDNTYSLELFOE 197

RESULT 6
US-10-634-905-5
; Sequence 5, Application US/10634905
; Publication No. US20040067225A1
; GENERAL INFORMATION:
; APPLICANT: GAN, Weinlu et al.
; TITLE OF INVENTION: ISOLATED HUMAN ENZYME PROTEINS, NUCLEIC
; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN ENZYME PROTEINS, AND USES
; TITLE OF INVENTION: THEREOF
; FILE REFERENCE: CL001147-DIV
; CURRENT APPLICATION NUMBER: US/10/634,905
; CURRENT FILING DATE: 2003-08-06
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5
; LENGTH: 197
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-634-905-5

Query Match          60.9%; Score 1039; DB 4; Length 197;
Best Local Similarity 100.0%; Pred. No. 7.7e-88;
Matches 197; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MKSALFTRFFILLPWILIVIMLDVTRRPVPLTPRPYFSPYAVGGRGRLPLRRGGPA 60
Db 1 MKSALFTRFFILLPWILIVIMLDVTRRPVPLTPRPYFSPYAVGGRGRLPLRRGGPA 60

Qy 61 HGTQKRNQSRPQPOPEQLPTIYAITYTSRPVQKAELTRANTFROVAQLHWILVEDAA 120
Db 61 HGTQKRNQSRPQPOPEQLPTIYAITYTSRPVQKAELTRANTFROVAQLHWILVEDAA 120
```

```
Qy 121 ARSELVSRFLARAGLPSTHLHVPTRRYKRPGLPRATEQRNAGLAWLRQRHQHORAQPGV 180
Db 121 ARSELVSRFLARAGLPSTHLHVPTRRYKRPGLPRATEQRNAGLAWLRQRHQHORAQPGV 180

Qy 181 LFFADDDNTYSLELFOE 197
Db 181 LFFADDDNTYSLELFOE 197

RESULT 7
US-10-698-190-16
; Sequence 16, Application US/10698190
; Publication No. US20040147469A1
; GENERAL INFORMATION:
; APPLICANT: Silver, J.
; APPLICANT: Grimpe, B.
; TITLE OF INVENTION: Methods of Inhibiting Glial Scar Formation
; FILE REFERENCE: CMRU-P01-018
; CURRENT APPLICATION NUMBER: US/10/698,190
; CURRENT FILING DATE: 2003-10-31
; NUMBER OF SEQ ID NOS: 40
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 16
; LENGTH: 335
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-698-190-16

Query Match          38.3%; Score 653; DB 4; Length 335;
Best Local Similarity 43.9%; Pred. No. 9e-52;
Matches 150; Conservative 41; Mismatches 93; Indels 58; Gaps 10;

Qy 1 MKSALFTRFFILLPWILIVIMLDVTRRPVPLTPRPYFSPYAVGGRGGA----- 50
Db 5 LKNVFLAYFLVLSIAGLYALVQLG-----QPCDCCLPPL-----RAAAEQLRQKD 48

Qy 51 -----RLPLRRGGPAHGTQKRNQSRPQPOPEQLPTIYAITYTSRPVQKAELTRAN 103
Db 49 LRISQLQAELELRPPPA-----PAQPPEALPTIVVTPTVARLVQKAEVLVSQ 98

Qy 104 TFRQVAQLHWILVEDAAARSELVSRFLARAGLPSTHLHVPTRP-----RYKRPGL--PRAT 157
Db 99 TLSLVPLRLHWLLVEDAEGPTPLVSLGAAASGLLFTHLVLTTPKAQRLEGEPEGVHPRGV 158

Qy 158 EQRNAGLAWLR-----ORHQHORAQPCVLFFADDDNTYSLELFOEMETTRKVSVMVPV 209
Db 159 EQRNKALDWLRGGGAVGGEKDPPTGQGVVTFADDDNTYSRELFEEMRWTRGVSVMPV 218

Qy 210 GLVGGRRYERPLVNGKVVGWYTGWRADRPFAIDMAGFAVSLQVILSNPKAVFKRRGSQP 269
Db 219 GLVGGRLRFEGPQVQDGRVVGFTAWEPSRPFVDMAGFAVALPLLLDKPNQAFDSTAPR- 277

Qy 270 GMOESDFLKOITTVESLEBKANNCTKVLVWHTTEKVNLANE 311
Db 278 GHLESSLHLVDPKDLEPRAANCTRLVWHTTEKPKMQE 319

RESULT 8
US-10-723-860-4238
; Sequence 4238, Application US/10723860
; Publication No. US20040253606A1
; GENERAL INFORMATION:
; APPLICANT: Aziz, Natasha
; APPLICANT: Ginsburg, Wendy M.
; APPLICANT: Zlotnik, Albert
; TITLE OF INVENTION: Methods of Diagnosis of Soft Tissue Sarcoma, Compositions &
; TITLE OF INVENTION: Methods for Screening for Soft Tissue Sarcoma Modulators
; FILE REFERENCE: 05882.0193.NPUS01
; CURRENT APPLICATION NUMBER: US/10/723,860
; CURRENT FILING DATE: 2003-11-26
; PRIOR FILING DATE: 2002-11-26
; NUMBER OF SEQ ID NOS: 8393
```

```
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 4238
; LENGTH: 335
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-723-860-4238

Query Match      38.3%; Score 653; DB 5; Length 335;
Best Local Similarity 43.9%; Pred. No. 9e-52;
Matches 150; Conservative 41; Mismatches 93; Indels 58; Gaps 10;

QY 1 MKSLFTRPILLPWILMLDVTDRP---VPPLTPRPSPYAVGGRGGA-----50
DB 5 LKVVFLAYFLVSTAGLLYALVQLG----QPCDCLPPL-----RAAEQLRQXD 48
QY 51 -----RLPLRRGGPARGTKRQNSRPQPEQ-LPTIYAITPTYSRPVQKAELTRLAN 103
DB 49 LRISQLQAELELRPPPA-----PAQPPPEALPTIYVVTPTVARLVQKAEVLRLSQ 98
QY 104 TFRVQAQLHWILVEDAAARSELVSRFLARAGLPSLHLVPTPR-----RYKRQGL--PRAT 157
DB 99 TLSLVPLRLHLLVEDAEGTPLVSGLLAASGLLFTLHLVLTPKAQLRREGEGFGVHPRGV 158
QY 158 EQRNAGLAWLR-----QRHQHQAQPGVLPFADDDNTYSLELFOEMRTTRKVSVPV 209
DB 159 EQRNKALDWLRGGRGAVGGEKPPPGTQGVVYFADDDNTYSRELFEEMRWTRGVSVPV 218
QY 210 GLVGGRRYERPLVGVGMYTGWADRPFAIDMAGFVLSQVILSNPKAVFKRGSQP 269
DB 219 GLVGLRFGEPQVQDGRVVGFTAWFSPRPFPVDMAGFVALPFLLLDKENAOFDSTAPR- 277
QY 270 GQESDFLQIITVELEPKANNCTKVLVWHTRTKVNLANE 311
DB 278 GHLESSLLHLVDPKOLEPRAANCTRLVWHTRTKPKMKQE 319

RESULT 9
US-11-097-143-13308
; Sequence 13308, Application US/11097143
; Publication No. US20050208558A1
; GENERAL INFORMATION:
; APPLICANT: Venter, J. Craig
; APPLICANT: et al.
; TITLE OF INVENTION: DETECTION KIT, SUCH AS NUCLEIC ACID
; TITLE OF INVENTION: ARRAYS, FOR DETECTING EXPRESSION OF 10,000 OR MORE
; FILE REFERENCE: CL000728
; CURRENT APPLICATION NUMBER: US/11/097,143
; PRIOR FILING DATE: 2005-04-04
; PRIOR FILING DATE: 1999-10-05
; PRIOR FILING DATE: 1999-10-19
; PRIOR FILING DATE: 1999-10-28
; PRIOR FILING DATE: 1999-11-12
; PRIOR FILING DATE: 1999-12-28
; PRIOR FILING DATE: 2000-01-12
; PRIOR FILING DATE: 2000-02-24
; PRIOR FILING DATE: 2000-03-23
; NUMBER OF SEQ ID NOS: 43008
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13308
; LENGTH: 479
; TYPE: PRT
; ORGANISM: DROSOPHILA
US-11-097-143-13308

Query Match      28.6%; Score 487.5; DB 6; Length 443;
Best Local Similarity 40.1%; Pred. No. 2.8e-36;
Matches 103; Conservative 48; Mismatches 91; Indels 15; Gaps 7;

QY 65 KENQSRPQPEP-POLPTIYAITPTYSRPVQKAELTRLANTPROVAQLHWILVEDAAARS 123
DB 149 RQFMQDQPSQSDYVQLPVIYFVTPTYPREQLPELTLAHTLLHLPRHLWLVADQEKCN 208
QY 124 ELVSRFLARAGLPSLHLVPTPRRYK--RPGLPRAEQNRNAGLAWLRQHQHQAQPGVL 181
DB 209 DYMDTLLYRFGMPFTHMVSMPESKERNKPA-PRGVANRRALQWIRQHN----LTNGIL 263

Query Match      29.5%; Score 502.5; DB 6; Length 479;
Best Local Similarity 40.0%; Pred. No. 1.3e-37;
Matches 102; Conservative 46; Mismatches 86; Indels 21; Gaps 6;

QY 80 PTIYAITPTYSRPVQKAELTRLANTPROVAQLHWILVEDAAARSELVSRFLARAGLPSLTH 139
DB 231 PLYIITPTYSRPEQLAELTRLGYTLKHVVNLLMLVIEDANKTNPLVGHITLDRIGVPVEY 290
QY 140 LHVPTPRRYK--RPGLPRAEQNRNAGLAWLRQHQHQAQPGVLPFADDDNTYSLELPOE 197
DB 291 MVAPMPEKYKTKAKPRGVSNNRGLGYLR-----EHATEGVLYFADDDNTYDISIFQ 345
QY 198 MTTRKVSVPVGLVGGRRYERPLVGVGMYTGWADRPFAIDMAGFVLSQVILSN 257
DB 346 MRYISKVAMWMPVGLVTKTGVSSPIIOAGKLVGYDGMIGGRKYPDVMAGFVSVKPKER 405
QY 258 PKAVFKRRGSGQMGESDFLQIITVE--ELPEPKANNCTKVLVWHTRTKTEK---VNLANEP 312
DB 406 PNA---QMPFKGYSGEDGFLRSFLAPLDDAEIELLADECEDEILTWHTQTKCNAPQAALNRT 462
QY 313 KY-----HLDTVKI 321
DB 463 RYKNTNLEHIDRLLV 477

RESULT 10
US-11-097-143-7506
; Sequence 7506, Application US/11097143
; Publication No. US20050208558A1
; GENERAL INFORMATION:
; APPLICANT: Venter, J. Craig
; APPLICANT: et al.
; TITLE OF INVENTION: DETECTION KIT, SUCH AS NUCLEIC ACID
; TITLE OF INVENTION: ARRAYS, FOR DETECTING EXPRESSION OF 10,000 OR MORE
; FILE REFERENCE: CL000728
; CURRENT APPLICATION NUMBER: US/11/097,143
; CURRENT FILING DATE: 2005-04-04
; PRIOR FILING DATE: 1999-10-05
; PRIOR FILING DATE: 1999-10-19
; PRIOR FILING DATE: 1999-10-28
; PRIOR FILING DATE: 1999-11-12
; PRIOR FILING DATE: 1999-12-28
; PRIOR FILING DATE: 2000-01-12
; PRIOR FILING DATE: 2000-02-24
; PRIOR FILING DATE: 2000-03-23
; NUMBER OF SEQ ID NOS: 43008
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7506
; LENGTH: 443
; TYPE: PRT
; ORGANISM: DROSOPHILA
US-11-097-143-7506

Query Match      28.6%; Score 487.5; DB 6; Length 443;
Best Local Similarity 40.1%; Pred. No. 2.8e-36;
Matches 103; Conservative 48; Mismatches 91; Indels 15; Gaps 7;

QY 65 KENQSRPQPEP-POLPTIYAITPTYSRPVQKAELTRLANTPROVAQLHWILVEDAAARS 123
DB 149 RQFMQDQPSQSDYVQLPVIYFVTPTYPREQLPELTLAHTLLHLPRHLWLVADQEKCN 208
QY 124 ELVSRFLARAGLPSLHLVPTPRRYK--RPGLPRAEQNRNAGLAWLRQHQHQAQPGVL 181
DB 209 DYMDTLLYRFGMPFTHMVSMPESKERNKPA-PRGVANRRALQWIRQHN----LTNGIL 263
```



Db 343 RWRGVSVMVPGVLGVLGVKQVQDSRVVGFHTAMEPNPFPVDVMDAGFAVALPILLAKP 402

QY 259 KAVFKRRGSGPQMQRSDFLKQITTVVEELSPKANNCTKVLV 298

Db 403 NAQFDSTAPRSHL-ESSLLSHLVDPKDLBPRAANCTRVLV 441

RESULT 14

US-10-403-571-34

; Sequence 34, Application US/10403571

; Publication No. US20040068763A1

; GENERAL INFORMATION:

; APPLICANT: Hopkins, Nancy

; APPLICANT: Golling, Gregory

; APPLICANT: Amsterdam, Adam

; APPLICANT: Sun, Zhoaxia

; TITLE OF INVENTION: Developmental Mutations in Zebrafish

; FILE REFERENCE: 01997/539002

; CURRENT APPLICATION NUMBER: US/10/403,571

; CURRENT FILING DATE: 2003-03-25

; PRIOR APPLICATION NUMBER: US 60/368,760

; PRIOR FILING DATE: 2002-03-29

; NUMBER OF SEQ ID NOS: 159

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 14

; LENGTH: 201

; TYPE: PRT

; ORGANISM: Danio rerio

US-10-403-571-34

Query Match 19.0%; Score 324.5; DB 4; Length 201;

Best Local Similarity 44.3%; Pred. No. 1.3e-21;

Matches 81; Conservative 24; Mismatches 45; Indels 33; Gaps 6;

QY 43 YAVRGCGARLPLRRGGPAHGTQKRNSRPQPOPEQO-----LPTIYAIT 86

Db 24 YALMQLGQRCDR-----DHEQSKDQIISQLKGELOKLQEHIKTSELSKKTDPRIYVIT 78

QY 87 PTYSRPVQKAELTRLANTFRQVLAHLVVEDAAARSELVSRFLARAGLPSTHLHVPTPR 146

Db 79 PTYARLVQKAELTRLSHTFLHVPQLHWIIVVEDAPQQTQLVSDFLSASGLTYTHLNKLTPK 138

QY 147 RYK-RPG-----LPRATEORNAGLWLRQ-----HCHQRA--QPGVLFEADDNDNTYSLEL 194

Db 139 ERKLOQGDNPWLKPRGAEQRNEGLRWNGSTVHGKEAAALEEAVVYFADDNDNTYSLOL 198

QY 195 FQE 197

Db 199 FEE 201

RESULT 15

US-10-437-963-174074

; Sequence 174074, Application US/10437963

; Publication No. US20040123343A1

; GENERAL INFORMATION:

; APPLICANT: La Rosa, Thomas J.

; APPLICANT: Kovalic, David K.

; APPLICANT: Zhou, Yihua

; APPLICANT: Cao, Yongwei

; APPLICANT: Wu, Wei

; APPLICANT: Boukharov, Andrey A.

; APPLICANT: Barbazuk, Brad

; APPLICANT: Li, Ping

; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With

; FILE REFERENCE: 38-21(53221)B

; CURRENT APPLICATION NUMBER: US/10/437,963

; CURRENT FILING DATE: 2003-05-14

; NUMBER OF SEQ ID NOS: 204966

; SEQ ID NO 174074

; LENGTH: 351

; TYPE: PRT

; ORGANISM: Oryza sativa

; FEATURE:

; OTHER INFORMATION: Clone ID: PAT\_MRT4530\_72050C.1.pep

US-10-437-963-174074

Query Match 18.3%; Score 312.5; DB 4; Length 351;

Best Local Similarity 28.7%; Pred. No. 3.4e-20;

Matches 93; Conservative 41; Mismatches 115; Indels 75; Gaps 12;

QY 23 LDVDTRRVPVPLTPRPYFSPYAVGRGGARLPLRRGGPAHGTQKRNQ-----SRPOPO 74

Db 37 LHLDLHLPLPP-----PATAILLVREDPPSVVVVDVDTLPA 71

QY 75 PEPQLPTIYAITPTYSRPVQKAELTRLANTFRQV-AQLHWILVEDAAARSELVSRFLARA 133

Db 72 AAEERKLLLVVTPTRARPLQAYTLRLAHTLRLAPSLLLLVVESGAATRD-TAALLRGC 130

QY 134 GLPSTHLHVPTP-----RRYKRPGLPRATEQRNAGLWLRQHRHQHQAQPGVLV 182

Db 131 GVMYRHLSSPVPDADQDPRRRGRQRDPVDSRARQRTALDHEHRLH-----GIVY 185

QY 183 FADDNDNTYSLELFOEMRTTRKVSVMVPG-LVGGRR---YERPLVENGKVGVGYTGWRA-- 236

Db 186 PAEDNVYSLDLFYHLRDIRSGFTWPVATLAPGSKTILQGEVCEGSRVVGWHTTDRSKN 245

QY 237 DRPFAIDMAGFAVSLQVILSNPKAVFKRGSQ-----PGMQESDFLKQITTV 284

Db 246 QRRFHVDMSGFAFN-----SSKLMDAKNRQHQAAMNYIRQLDTAKEGFOETAFIEQLVEDE 300

QY 285 -ELEPKANNCTKVLVWHTRTTEKVN 307

Db 301 THMEGVPPGCCSKIMNFHLHLEDKN 324

Search completed: May 9, 2006, 10:58:56

Job time : 166 secs

GenCore version 5.1.8  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: May 9, 2006, 10:55:30 ; Search time 18 Seconds  
(without alignments)  
830.554 Million cell updates/sec

Title: US-10-634-905-2  
Perfect score: 1705  
Sequence: 1 MKSALFTRFFILLPWILVI.....EKVLANEPKYHLDTVKIEV 323

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 235405 seqs, 46284737 residues

Total number of hits satisfying chosen parameters: 235405

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published Applications AA New:\*

- 1: /SIDSS/ptodata/1/pubpaa/US08\_NEW\_PUB.pep.\*
- 2: /SIDSS/ptodata/1/pubpaa/US06\_NEW\_PUB.pep.\*
- 3: /SIDSS/ptodata/1/pubpaa/US07\_NEW\_PUB.pep.\*
- 4: /SIDSS/ptodata/1/pubpaa/US08\_NEW\_PUB.pep.\*
- 5: /SIDSS/ptodata/1/pubpaa/US09\_NEW\_PUB.pep.\*
- 6: /SIDSS/ptodata/1/pubpaa/US09\_NEW\_PUB.pep.\*
- 7: /SIDSS/ptodata/1/pubpaa/US09\_NEW\_PUB.pep.\*
- 8: /SIDSS/ptodata/1/pubpaa/US10\_NEW\_PUB.pep.\*
- 9: /SIDSS/ptodata/1/pubpaa/US10\_NEW\_PUB.pep.\*
- 10: /SIDSS/ptodata/1/pubpaa/US11\_NEW\_PUB.pep.\*
- 11: /SIDSS/ptodata/1/pubpaa/US11\_NEW\_PUB.pep.\*
- 12: /SIDSS/ptodata/1/pubpaa/US60\_NEW\_PUB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1705	100.0	323	9	US-10-330-773-566
2	1207	70.8	290	9	US-10-330-773-563
3	249	14.6	408	11	US-11-096-568A-18021
4	249	14.6	440	11	US-11-096-568A-18020
5	249	14.6	448	11	US-11-096-568A-18019
6	243.5	14.3	354	11	US-11-096-568A-2084
7	243.5	14.3	385	11	US-11-096-568A-2083
8	243.5	14.3	441	11	US-11-096-568A-2082
9	193.5	11.3	492	9	US-10-467-962B-51
10	117.5	6.9	96	9	US-10-475-075-809
11	99.5	5.8	813	8	US-10-511-814-16
12	99.5	5.8	813	9	US-10-661-966-19
13	95	5.6	424	11	US-11-096-568A-22088
14	94.5	5.5	296	11	US-11-096-568A-1548
15	94.5	5.5	392	11	US-11-096-568A-21010
16	94	5.5	527	11	US-11-096-568A-23555
17	93	5.5	783	8	US-10-509-131-25
18	93	5.5	783	11	US-11-186-284-59
19	92.5	5.4	721	11	US-11-186-284-67
20	92	5.4	480	9	US-10-330-773-786
21	90.5	5.3	278	11	US-11-096-568A-9264

22	90.5	5.3	278	11	US-11-096-568A-9266	Sequence 9266, Ap
23	90.5	5.3	319	11	US-11-096-568A-9263	Sequence 9263, Ap
24	90	5.3	254	11	US-11-096-568A-3889	Sequence 3889, Ap
25	90	5.3	293	11	US-11-096-568A-3888	Sequence 3888, Ap
26	89	5.2	461	11	US-11-096-568A-18013	Sequence 18013, A
27	89	5.2	462	11	US-11-096-568A-18012	Sequence 18012, A
28	89	5.2	463	11	US-11-096-568A-18011	Sequence 18011, A
29	89	5.2	614	11	US-11-015-546A-20	Sequence 20, Appl
30	88	5.2	221	11	US-11-096-568A-17196	Sequence 17196, A
31	88	5.2	233	11	US-11-096-568A-17195	Sequence 17195, A
32	88	5.2	1189	11	US-11-182-885-3	Sequence 3, Appli
33	87.5	5.1	250	11	US-11-096-568A-21169	Sequence 21169, A
34	87.5	5.1	409	11	US-11-096-568A-9548	Sequence 9548, Ap
35	87	5.1	416	11	US-11-096-568A-16843	Sequence 16843, A
36	87	5.1	528	11	US-11-087-099-7839	Sequence 7839, Ap
37	86.5	5.1	663	11	US-11-072-512-3771	Sequence 3771, Ap
38	86.5	5.1	1052	11	US-11-020-602-208	Sequence 208, App
39	86.5	5.1	2725	11	US-11-113-424-52	Sequence 52, Appl
40	86.5	5.1	2725	11	US-11-100-640-10	Sequence 10, Appl
41	86.5	5.1	2725	11	US-11-100-640-16	Sequence 16, Appl
42	86	5.0	534	11	US-11-096-568A-24457	Sequence 24457, A
43	86	5.0	548	11	US-11-096-568A-24456	Sequence 24456, A
44	85.5	5.0	390	11	US-11-096-568A-21913	Sequence 21913, A
45	85	5.0	281	11	US-11-096-568A-22579	Sequence 22579, A

ALIGNMENTS

RESULT 1  
US-10-330-773-566  
; Sequence 566, Application US/10330773  
; Publication No. US20060040262A1  
; GENERAL INFORMATION:  
; APPLICANT: David W. Morris  
; APPLICANT: Marc Malandro  
; TITLE OF INVENTION: Novel Compositions and Methods in Cancer  
; FILE REFERENCE: 529452001300  
; CURRENT APPLICATION NUMBER: US/10/330,773  
; CURRENT FILING DATE: 2002-12-27  
; NUMBER OF SEQ ID NOS: 981  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 566  
; LENGTH: 323  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-330-773-566

Query Match	100.0%;	Score 1705;	DB 9;	Length 323;
Best Local Similarity	100.0%;	Pred. No. 1.9e-148;		
Matches 323;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
Qy	1	MKSALFTRFFILLPWILVIIMLDVTRRPVPLTRPPSPYAVGGRGARLPLRRGGPA	60	
Db	1	MKSALFTRFFILLPWILVIIMLDVTRRPVPLTRPPSPYAVGGRGARLPLRRGGPA	60	
Qy	61	HGTQKNQSRPQPQPPQPTTIVAITPTYSRPVQKAELTRLANTFRQVAGLHWLVEDAA	120	
Db	61	HGTQKNQSRPQPQPPQPTTIVAITPTYSRPVQKAELTRLANTFRQVAGLHWLVEDAA	120	
Qy	121	ARSELVSRFLARAGLPSLHVPTPRRYKRPGLPRATEQRNAGLMLRQHQHQAQPGV	180	
Db	121	ARSELVSRFLARAGLPSLHVPTPRRYKRPGLPRATEQRNAGLMLRQHQHQAQPGV	180	
Qy	181	LPFADDNTYLSLFOEMRTTRKVSVPVGLVCGRBYERPLVNGKVGVHYTGWRDRPF	240	
Db	181	LPFADDNTYLSLFOEMRTTRKVSVPVGLVCGRBYERPLVNGKVGVHYTGWRDRPF	240	
Qy	241	AIDMAGFVSLQVILSNPKAVFKRGSCQPMQSSDFLKQITTVVEELEPKANNCTKVLVWH	300	
Db	241	AIDMAGFVSLQVILSNPKAVFKRGSCQPMQSSDFLKQITTVVEELEPKANNCTKVLVWH	300	
Qy	301	TRTEKVNLANEPKYHLDTVKIEV	323	

```
|||||
Db 301 TRTEKVNLANEPKYHLDTVKIEV 323

RESULT 2
US-10-330-773-563
; Sequence 563, Application US/10330773
; Publication No. US20060040282A1
; GENERAL INFORMATION:
; APPLICANT: David W. Morris
; APPLICANT: Marc Malandro
; TITLE OF INVENTION: Novel Compositions and Methods in Cancer
; FILE REFERENCE: 529452001300
; CURRENT APPLICATION NUMBER: US/10/330, 773
; CURRENT FILING DATE: 2002-12-27
; NUMBER OF SEQ ID NOS: 981
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 563
; LENGTH: 290
; TYPE: PRT
; ORGANISM: Mus musculus
US-10-330-773-563

Query Match 70.8%; Score 1207; DB 9; Length 290;
Best Local Similarity 73.8%; Pred. No. 6.8e-103;
Matches 239; Conservative 8; Mismatches 17; Indels 60; Gaps 3;

QY 1 MSALETRFILLPWILVIIMLDVDRPVPPLTPRPYFSPYAVGRRGGARLPRLRRGPA 60
||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
Db 26 MSALECSRFPILLPWILVIIMLDVDRPAPQLTSRPYFSPHVAVCGGSRVPLRRSPG 85
||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
QY 61 H-GTQKRNQSRPQPEPQLTIYAITPTYSRVPVQKAELTRLANTRPQVAQLHWILVEDA 119
:||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
Db 86 RDAAEKRNESRPOLEPEPLTIYAITPTYSRP- 118
||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
QY 120 AARSELVSRFLARAGLPSTHLVPTPRYKRKGLPRATEQRNAGLAWLRQRHQHQAOPG 179
||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
Db 119 -----RPLPRATEQRNAGLAWLRQRHQHQAOPG 148
||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
QY 180 VLFFADDDNTYSLELFQEMRTTRKVSVPVGLVGGRRYERPLVNGKVVGWYTGWRADRP 239
||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
Db 149 VLFFADDDNTYSLELFQEMRTTRKVSVPVGLVGGRRYERPLVNGKVVGWYTGWRADRP 208
||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
QY 240 FAIDMAGFAVLSQVILSNKAPVKRRGSGQMOESDFLKQITTVEELEPKANNCTKVLVW 299
||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
Db 209 FAIDMAGFAVLSQVILSNKAPVKRRGSGQMOESDFLKQITTVEELEPK--NCTKVLVW 266
||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
QY 300 HTRTEKVNLANEPKYHLDTVKIEV 323
||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
Db 267 HTRTEKVNLANEPKYHLDTVNIEV 290
||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||

RESULT 3
US-11-096-568A-18021
; Sequence 18021, Application US/11096568A
; Publication No. US20060048240A1
; GENERAL INFORMATION:
; APPLICANT: Alexandrov, Nikolai et al.
; TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypeptides
; FILE REFERENCE: 2750-1592PUS2
; CURRENT APPLICATION NUMBER: US/11/096, 568A
; CURRENT FILING DATE: 2005-04-01
; NUMBER OF SEQ ID NOS: 34471
; SEQ ID NO 18021
; LENGTH: 408
; TYPE: PRT
; ORGANISM: Zea mays subsp. mays
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)..(408)
; OTHER INFORMATION: Ceres Seq. ID no. 12362396
US-11-096-568A-18021
```

```
Query Match 14.6%; Score 249; DB 11; Length 408;
Best Local Similarity 30.5%; Pred. No. 5.6e-15;
Matches 78; Conservative 44; Mismatches 102; Indels 32; Gaps 11;

QY 85 ITPTYSRVPVQKAELTRLANTRPROV-AQLHWILVEDAAARSELVSRFLARAGLPSTHLHVP 143
:||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
Db 160 VTTSVVRPHQAYYLNRLAHVLKNVPPPLLIWVAE-WPQSRETAAILRSSGVMYRHLCN 218
||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
QY 144 TPRRYKRPLPRATEQRNAGLAWLRQRHQHQAOPGVLFADDDNTYSLELFQEMRTTRK 203
:||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
Db 219 RNSTNIRKII--VCQKNAIF-----HIKKHRLDGIHVHFADEERVYSVDLFEDMRKIRR 270
:||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
QY 204 VSWVPGLVGGRRY-----ERPLVNGKVVGWYTGWRADRP--FAIDMAGFAVLSQVILSN 257
||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
Db 271 FGTWVPVATHVGARYKVLGEGPVCRCGNQVTGMHTNKRGGVPRRFPFIGSGFAFN-STILWD 329
||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
QY 258 P-----KAVFKRRGSGQPGMOESDFLKQITTVS-ELEPKANNCTKVLVHTRTEKVN 308
:||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
Db 330 PORWNSPTLESILHSGGRRGLQESRFIEKVEDETQMEGLADNCTRVVMWNLDPPL 389
||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
QY 309 ANEP-----KYHLDTV 319
||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
Db 390 -NYPTSWOLQKNLDIV 404
||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||

RESULT 4
US-11-096-568A-18020
; Sequence 18020, Application US/11096568A
; Publication No. US20060048240A1
; GENERAL INFORMATION:
; APPLICANT: Alexandrov, Nikolai et al.
; TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypeptides
; FILE REFERENCE: 2750-1592PUS2
; CURRENT APPLICATION NUMBER: US/11/096, 568A
; CURRENT FILING DATE: 2005-04-01
; NUMBER OF SEQ ID NOS: 34471
; SEQ ID NO 18020
; LENGTH: 440
; TYPE: PRT
; ORGANISM: Zea mays subsp. mays
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)..(440)
; OTHER INFORMATION: Ceres Seq. ID no. 12362395
US-11-096-568A-18020

Query Match 14.6%; Score 249; DB 11; Length 440;
Best Local Similarity 30.5%; Pred. No. 6.2e-15;
Matches 78; Conservative 44; Mismatches 102; Indels 32; Gaps 11;

QY 85 ITPTYSRVPVQKAELTRLANTRPROV-AQLHWILVEDAAARSELVSRFLARAGLPSTHLHVP 143
:||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
Db 192 VTTSVVRPHQAYYLNRLAHVLKNVPPPLLIWVAE-WPQSRETAAILRSSGVMYRHLCN 250
||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
QY 144 TPRRYKRPLPRATEQRNAGLAWLRQRHQHQAOPGVLFADDDNTYSLELFQEMRTTRK 203
:||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
Db 251 RNSTNIRKII--VCQKNAIF-----HIKKHRLDGIHVHFADEERVYSVDLFEDMRKIRR 302
:||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
QY 204 VSWVPGLVGGRRY-----ERPLVNGKVVGWYTGWRADRP--FAIDMAGFAVLSQVILSN 257
||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
Db 303 FGTWVPVATHVGARYKVLGEGPVCRCGNQVTGMHTNKRGGVPRRFPFIGSGFAFN-STILWD 361
||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
QY 258 P-----KAVFKRRGSGQPGMOESDFLKQITTVS-ELEPKANNCTKVLVHTRTEKVN 308
:||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
Db 362 PORWNSPTLESILHSGGRRGLQESRFIEKVEDETQMEGLADNCTRVVMWNLDPPL 421
||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
QY 309 ANEP-----KYHLDTV 319
||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
Db 422 -NYPTSWOLQKNLDIV 436
||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
```

RESULT 5  
US-11-096-568A-18019  
; Sequence 18019, Application US/11096568A  
; Publication No. US20060048240A1  
; GENERAL INFORMATION:  
; APPLICANT: Alexandrov, Nikolai et al.  
; TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypeptides  
; FILE REFERENCE: 2750-1592PUS2  
; CURRENT APPLICATION NUMBER: US/11/096.568A  
; CURRENT FILING DATE: 2005-04-01  
; NUMBER OF SEQ ID NOS: 34471  
; SEQ ID NO 18019  
; LENGTH: 448  
; TYPE: PRT  
; ORGANISM: Zea mays subsp. mays  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (1)..(448)  
; OTHER INFORMATION: Ceres Seq. ID no. 12362394  
US-11-096-568A-18019

Query Match 14.6%; Score 249; DB 11; Length 448;  
Best Local Similarity 30.5%; Pred. No. 6.3e-15;  
Matches 78; Conservative 44; Mismatches 102; Indels 32; Gaps 11;  
QY 85 ITPTYSRPVQKAEI...AQLHWILVEDAAARSELVSRFLA 143  
DB 200 VTTTSVPQAYLN...LAVHKNVPPPELLVAE-WPYQRETAIILRSSGVMYRHLICN 258  
QY 144 TPRYKRPGLPRATE...QORNAGLAWLRQHQHQAQPGVLPFADDDNTYSLSLEFQEMRTTRK 203  
DB 259 RNSTNIRKII--VCQKNAIF-----HIKHLRDGIVHFADEERVYSVDLFEDMRKIR 310  
QY 204 VSWPVLGVGRRY----ERPLVGVGVGWTGWRADRP--FAIDMAGFAVSLQVILSN 257  
DB 311 FGTWPVATHGARYKV...VLEGVPCVCRGNQVTGHTNKRGGVPRRPIGFSGAFN-STILLWD 369  
QY 258 P-----KAVFKRGSQPMQESDFLKQITTTVE-ELEPKANNCTKVLVWHTRTEKVN 308  
DB 370 PQWNSPTLESII...HSGRGLQESRFIEKLVEDETOEGLADNCTRVMWNLDLEPPQL 429  
QY 309 ANEP-----KYHLDTV 319  
DB 430 -NYPTSWQLQKNDIV 444

RESULT 6  
US-11-096-568A-2084  
; Sequence 2084, Application US/11096568A  
; Publication No. US20060048240A1  
; GENERAL INFORMATION:  
; APPLICANT: Alexandrov, Nikolai et al.  
; TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypeptides  
; FILE REFERENCE: 2750-1592PUS2  
; CURRENT APPLICATION NUMBER: US/11/096.568A  
; CURRENT FILING DATE: 2005-04-01  
; NUMBER OF SEQ ID NOS: 34471  
; SEQ ID NO 2084  
; LENGTH: 354  
; TYPE: PRT  
; ORGANISM: Zea mays subsp. mays  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (1)..(354)  
; OTHER INFORMATION: Ceres Seq. ID no. 15180705  
US-11-096-568A-2084

Query Match 14.3%; Score 243.5; DB 11; Length 354;  
Best Local Similarity 28.9%; Pred. No. 1.5e-14;  
Matches 76; Conservative 52; Mismatches 94; Indels 41; Gaps 12;

QY 76 EPOLPT---IYAITPTYSRPVQKAEI...AQLHWILVEDAAARSELVSRFLA 131  
DB 91 ESDIPVKLLIIVTITS...VRPQAYVYLNRLAHTLKAQAPLLLWVE-WPEQSYETABILR 149  
QY 132 RAGLPSTHL-----HVPTPRYKRPGLPRATE...QORNAGLAWLRQHQHQAQPGVLPFADDD 187  
DB 150 SSGVMYRHLMKCRKNTTSVSRKI-----AVCQRNNAIYHVKRHH-----LDGIMHFADEE 197  
QY 188 NTYSLSLEFQEMRTTRK...VSWPVLGVGRRY----ERPLVGVGVGWTGWRAD-----R 238  
DB 198 RSYADVFEEQKIRRF...GWSWPVTIHGTYKRAVLEGPICANRVMRWHITVTAQKKSSTR 257  
QY 239 PFAIDMAGFAVSLQVILSNPK-----AVFKRGSQPMQESDFLKQITTTVE-ELEPK 289  
DB 258 RPIGFSAFAN-STILLWD...PQRWNRPPMDSVI VHTGGRGLQESRFIEKLVKSERQIEGL 316  
QY 290 ANNCTKVLVWHTRTEKVNLANEP 312  
DB 317 PDNCNRVMVMNFNLEPPQL-NYP 338  
RESULT 7  
US-11-096-568A-2083  
; Sequence 2083, Application US/11096568A  
; Publication No. US20060048240A1  
; GENERAL INFORMATION:  
; APPLICANT: Alexandrov, Nikolai et al.  
; TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypeptides  
; FILE REFERENCE: 2750-1592PUS2  
; CURRENT APPLICATION NUMBER: US/11/096.568A  
; CURRENT FILING DATE: 2005-04-01  
; NUMBER OF SEQ ID NOS: 34471  
; SEQ ID NO 2083  
; LENGTH: 385  
; TYPE: PRT  
; ORGANISM: Zea mays subsp. mays  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (1)..(385)  
; OTHER INFORMATION: Ceres Seq. ID no. 15180704  
US-11-096-568A-2083

Query Match 14.3%; Score 243.5; DB 11; Length 385;  
Best Local Similarity 28.9%; Pred. No. 1.7e-14;  
Matches 76; Conservative 52; Mismatches 94; Indels 41; Gaps 12;  
QY 76 EPOLPT---IYAITPTYSRPVQKAEI...AQLHWILVEDAAARSELVSRFLA 131  
DB 122 ESDIPVKLLIIVTITS...VRPQAYVYLNRLAHTLKAQAPLLLWVE-WPEQSYETABILR 180  
QY 132 RAGLPSTHL-----HVPTPRYKRPGLPRATE...QORNAGLAWLRQHQHQAQPGVLPFADDD 187  
DB 181 SSGVMYRHLMKCRKNTTSVSRKI-----AVCQRNNAIYHVKRHH-----LDGIMHFADEE 228  
QY 188 NTYSLSLEFQEMRTTRK...VSWPVLGVGRRY----ERPLVGVGVGWTGWRAD-----R 238  
DB 229 RSYADVFEEQKIRRF...GWSWPVTIHGTYKRAVLEGPICANRVMRWHITVTAQKKSSTR 288  
QY 239 PFAIDMAGFAVSLQVILSNPK-----AVFKRGSQPMQESDFLKQITTTVE-ELEPK 289  
DB 289 RPIGFSAFAN-STILLWD...PQRWNRPPMDSVI VHTGGRGLQESRFIEKLVKSERQIEGL 347  
QY 290 ANNCTKVLVWHTRTEKVNLANEP 312  
DB 348 PDNCNRVMVMNFNLEPPQL-NYP 369

RESULT 8  
US-11-096-568A-2082  
; Sequence 2082, Application US/11096568A  
; Publication No. US20060048240A1  
; GENERAL INFORMATION:

APPLICANT: Alexandrov, Nikolai et al.  
TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypeptides  
FILE OF INVENTION: Therby  
FILE REFERENCE: 2750-1592PUS2  
CURRENT APPLICATION NUMBER: US/11/096,568A  
CURRENT FILING DATE: 2005-04-01  
NUMBER OF SEQ ID NOS: 34471  
SEQ ID NO 2082  
LENGTH: 441  
TYPE: PRT  
ORGANISM: Zea mays subsp. mays  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (1)..(441)  
OTHER INFORMATION: Ceres Seq. ID no. 15180703  
US-11-096-568A-2082

Query Match 14.3%; Score 243.5; DB 11; Length 441;  
Best Local Similarity 28.9%; Pred. No. 2e-14; Mismatches 94; Indels 41; Gaps 12;  
Matches 76; Conservative 52;  
QY 76 EPQLPT---IYAITPTYSRPVQKAEI...AQLHWILVEDAAARSELVSRFLA 131  
DB 178 ESDIPVKLLIIVTITSVRPQQAYLNRLAHTLKAVQAPLLMLVE-WPEQSVETAEILR 236  
QY 132 RAGLPSTHL-----HVTTPRYKRGPLPRATEQORNAGLAWLRQSHQHORAQPGVLPFADDD 187  
DB 237 SSGVMYRHLCKRNTTTSVRKI-----AVCQRNNAIYVGRHH-----LDGIMHFADSE 284  
QY 188 NTYSLELFOEMRTTRKVSVPVGLVGRRY---ERPLVGVKVGWYTGWRAD-----R 238  
DB 285 RYSADVFEEMQIRKRGSPWVTIHGTYKRAVLEGPICAKRVMRWHTVQTAQKSSSTR 344  
QY 239 PFAIDMAGFAVSLQVILSNPK-----AVFKRGSQPGMQESDFLKQITVE-ELEPK 289  
DB 345 RPIPGSAFAFN-STLLMDPQRNRPMDSVIHTGGRGGLQESRFIEKLKVKSERQIEGL 403  
QY 290 ANNCTKVLVWHTTEKVNLANEP 312  
DB 404 PDCNRVMVNFNLEPPQL-NYP 425

RESULT 9  
US-10-467-962B-51  
Sequence 51, Application US/10467962B  
Publication No. US20050246784A1  
GENERAL INFORMATION:  
APPLICANT: Plesch, Gunnar  
APPLICANT: Blau, Astrid  
APPLICANT: Daeschner, Klaus  
APPLICANT: Klein, Mathieu  
TITLE OF INVENTION: Identification of Herbicidally Active Substances  
FILE REFERENCE: 2000 857  
CURRENT APPLICATION NUMBER: US/10/467,962B  
CURRENT FILING DATE: 2003-08-14  
PRIOR APPLICATION NUMBER: PCT/EP02/01466  
PRIOR FILING DATE: 2002-02-13  
NUMBER OF SEQ ID NOS: 109  
SOFTWARE: PatentIn Vers. 2.0  
SEQ ID NO 51  
LENGTH: 492  
TYPE: PRT  
ORGANISM: Arabidopsis thaliana  
US-10-467-962B-51

Query Match 11.3%; Score 193.5; DB 9; Length 492;  
Best Local Similarity 22.2%; Pred. No. 8.6e-10; Mismatches 55; Indels 119; Gaps 16;  
Matches 81; Conservative 55;  
QY 39 YFSPYAVRGGARLP...AHGTQKRN-----QSRPQPPPEP----- 77  
DB 73 YTTFFLFAGNGGVSQ...LRLKPLETATNSTVKKNSRVVVGHHGIRIRPWPHPNPVLELRAH 132

QY 78 QL-----PTIYAITPTYSRPVQKAEI...AQLHWILVEDA 119  
DB 133 QLLVRVQKEQKSNYGVPRPRIVVVTYVTRFTQALHTGVNHSMLVPDYDLVWIVVEAG 192  
QY 120 AARSELSRFLARAGLPSTHLHVPTRPRYKRGPLPRATEQORNAGLAWLR---QRHQHORA 176  
DB 193 GITNETAS-FIAKSGLKTIHLGFDQK-----MPTWEDRHKLETKMRLHALRVVREKK 244  
QY 177 QPGVLFPFADDDNTYSLELFOEMRTTRKVSVPVGLVG-----GRRYERP 220  
DB 245 LDGIVWFADDSNMHSMELFDEIQTVKFGALSVGILAHSGNADELSSILKNEQGNKEKP 304  
QY 221 LV-----ENGKVVGWY-----TGWADR-----PFAIDMAGFAVSLQVILSN 257  
DB 305 SMPIQGPCNSSEKLVGWHIFNTQPYAKTAVYIDEKAPVMPSPKMEWSGFVL-----N 357  
QY 258 PKAVFKRR-GSQPGMQESDFLKQITVEE-----LEPKANNCTKVLVW 299  
DB 358 SELLWKESLDDKPA-----WVKDLSLLDDGYAEIESPLSLVKDPSMVEPLGSGRRVLLW 412  
QY 300 HTRTE 304  
DB 413 WLURVE 417

RESULT 10  
US-10-475-075-809  
Sequence 809, Application US/10475075  
Publication No. US20060053498A1  
GENERAL INFORMATION:  
APPLICANT: Bejanin, Stephane  
APPLICANT: Tanaka, Hiroaki  
APPLICANT: Dumas Milne Edwards, Jean-Baptiste  
APPLICANT: Jobert, Severin  
APPLICANT: Giordano, Jean-Yves  
TITLE OF INVENTION: Full-length human cDNAs encoding potentially secreted proteins  
FILE REFERENCE: G-081US03PCT  
CURRENT APPLICATION NUMBER: US/10/475,075  
CURRENT FILING DATE: 2003-10-17  
PRIOR APPLICATION NUMBER: PCT/IB01/00914  
PRIOR FILING DATE: 2001-04-18  
NUMBER OF SEQ ID NOS: 918  
SOFTWARE: Patent.pm  
SEQ ID NO 809  
LENGTH: 96  
TYPE: PRT  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: SIGNAL  
LOCATION: -33...-1  
US-10-475-075-809

Query Match 6.9%; Score 117.5; DB 9; Length 96;  
Best Local Similarity 31.4%; Pred. No. 0.00095; Mismatches 37; Conservative 13; Mismatches 25; Indels 43; Gaps 6;  
QY 1 MKSALFTRPFILLPWILVIMLDVDTTRP---VPPLTRPYFSPYAVRGGA----- 50  
DB 5 LKNVFLAYFLVSIAGLLYALVOLG---QPCDCLPPL-----RAAAEQLRQKD 48  
QY 51 -----RLPLRRGGPAHGTCQENQSRPQPBPQ-LPTIYAITPTYSRPVQKAEI... 101  
DB 49 LRISQLQAEILRRPPPA-----PAQPPEALPTIYVVTYARLVQKAEILVRL 96

RESULT 11  
US-10-511-814-16  
Sequence 16, Application US/10511814  
Publication No. US20060088472A1  
GENERAL INFORMATION:  
APPLICANT: McCance, Dennis  
APPLICANT: Westbrook, III, Thomas F.  
TITLE OF INVENTION: E7 REGULATION OF P21 (CIP1) THROUGH AKT



```
; FILE REFERENCE: 21108.0016U2
; CURRENT APPLICATION NUMBER: US/10/511.814
; PRIOR FILING DATE: 2004-10-19
; PRIOR FILING DATE: 2004-10-19
; PRIOR FILING DATE: 2003-04-21
; PRIOR FILING DATE: 2002-04-19
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 16
; LENGTH: 813
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence./Note =
; OTHER INFORMATION: Synthetic Construct
US-10-511-814-16

Query Match          5.8%; Score 99.5; DB 8; Length 813;
Best Local Similarity 21.4%; Pred. No. 0.66;
Matches 55; Conservative 43; Mismatches 102; Indels 57; Gaps 15;

Qy 25 VDTRRPVP-----LTPRPVPSYAVGRGARGARLPRLRGGAHGT----QKKNQSRPQ 72
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 458 LDAQRPRPQKPHHEDWEILPNEFIQYKVGSGSGFTVYR--GEFFGTVAIKKLNVDPT 515
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 73 PQPEPQLPTIYAITPTYSRPVQKAELTRLANTFRQVAQLHWILVEDAAARSELVSRFLAR 132
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 516 PSQ-----MAAFKNEVAVLKTRHLNV---LLFMGWVREPEIA-----IITQCEG 558
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 133 AGLPSTHLHVPTPR-RYKRPGLPRATEQRNAGLAWLRQRH-QHORAQPGVLFADDDNTY 190
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 559 SSL-YRHHVQEPREVFEFGAIIIDLKQVSLGMYLHSHKNIHRLDLTKNNIFLMDDMSTV 617
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 191 SLELFQEMRTTRKVSVPVGLVGRRYERPLVNGKVVGYTGWRA-----DRPFA- 241
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 618 KIGDFGLATVKTK---WTNV--GGQQQQQP---TGSIL-----WMAPEVIRMQDDNPYP 664
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 242 -IDMAGFAVSLQVILSN 257
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 665 QSDVYSFGICMYEILSS 681

RESULT 13
US-11-096-568A-22088
; Sequence 22088, Application US/11096568A
; Publication No. US20060048240A1
; GENERAL INFORMATION:
; APPLICANT: Alexandrov, Nikolai et al.
; TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypeptides
; FILE REFERENCE: 2750-1592PUS2
; CURRENT APPLICATION NUMBER: US/11/096,568A
; CURRENT FILING DATE: 2005-04-01
; NUMBER OF SEQ ID NOS: 34471
; SEQ ID NO 22088
; LENGTH: 424
; TYPE: PRT
; ORGANISM: Zea mays subsp. mays
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)-(424)
; OTHER INFORMATION: Ceres Seq. ID no. 12407511
US-11-096-568A-22088

Query Match          5.6%; Score 95; DB 11; Length 424;
Best Local Similarity 25.7%; Pred. No. 0.74;
Matches 54; Conservative 10; Mismatches 60; Indels 86; Gaps 10;

Qy 29 RVPVPLTPR-PYFSPYAVG-----RGGARLP-----LRRGGPAHGTQKRNQSR 70
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 98 RPLPPLRSRGPAPETVRAAAARGAVRGARPPAPAPVQAGRGRAQREARGGHQP 157
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 71 PQPQEPQLPTIYAITPTYSRPVQKAELTRLANTFRQVAQLHWILVEDAAARSELVSRFL 130
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 158 PQRAPPP-----RPAQAERRRADG-----GALHHL---RRASPRGD----- 191
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 131 ARAGLPSTHLHVPTPRRYKRPGLPRATEQRNAGLAWLRQRHQRHORAQPGVLFADDDNTY 190
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 192 --AGL-----RGAPRAPDAR--GGAGARGGHPVAHQDPGVLPRAGQ---- 228
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 191 SLELFQEMRTTRKVSVPVGLVGRRYERP 220
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 229 -----ARAGGARHARP 239

RESULT 14
US-11-096-568A-1548
; Sequence 1548, Application US/11096568A
; Publication No. US20060048240A1
; GENERAL INFORMATION:
; APPLICANT: Alexandrov, Nikolai et al.
; TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypeptides
; FILE REFERENCE: 2750-1592PUS2
; CURRENT APPLICATION NUMBER: US/11/096,568A
```

```
; FILE REFERENCE: 21108.0016U2
; CURRENT APPLICATION NUMBER: US/10/511.814
; PRIOR FILING DATE: 2004-10-19
; PRIOR FILING DATE: 2004-10-19
; PRIOR FILING DATE: 2003-04-21
; PRIOR FILING DATE: 2002-04-19
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 16
; LENGTH: 813
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence./Note =
; OTHER INFORMATION: Synthetic Construct
US-10-511-814-16

Query Match          5.8%; Score 99.5; DB 8; Length 813;
Best Local Similarity 21.4%; Pred. No. 0.66;
Matches 55; Conservative 43; Mismatches 102; Indels 57; Gaps 15;

Qy 25 VDTRRPVP-----LTPRPVPSYAVGRGARGARLPRLRGGAHGT----QKKNQSRPQ 72
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 458 LDAQRPRPQKPHHEDWEILPNEFIQYKVGSGSGFTVYR--GEFFGTVAIKKLNVDPT 515
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 73 PQPEPQLPTIYAITPTYSRPVQKAELTRLANTFRQVAQLHWILVEDAAARSELVSRFLAR 132
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 516 PSQ-----MAAFKNEVAVLKTRHLNV---LLFMGWVREPEIA-----IITQCEG 558
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 133 AGLPSTHLHVPTPR-RYKRPGLPRATEQRNAGLAWLRQRH-QHORAQPGVLFADDDNTY 190
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 559 SSL-YRHHVQEPREVFEFGAIIIDLKQVSLGMYLHSHKNIHRLDLTKNNIFLMDDMSTV 617
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 191 SLELFQEMRTTRKVSVPVGLVGRRYERPLVNGKVVGYTGWRA-----DRPFA- 241
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 618 KIGDFGLATVKTK---WTNV--GGQQQQQP---TGSIL-----WMAPEVIRMQDDNPYP 664
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 242 -IDMAGFAVSLQVILSN 257
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 665 QSDVYSFGICMYEILSS 681

RESULT 12
US-10-661-966-19
; Sequence 19, Application US/10661966
; Publication No. US20050277118A1
; GENERAL INFORMATION:
; APPLICANT: Nelson, Matthew Roberts
; APPLICANT: Braun, Andreas
; TITLE OF INVENTION: METHODS FOR IDENTIFYING SUBJECTS AT RISK
; FILE REFERENCE: 524592003800
; CURRENT APPLICATION NUMBER: US/10/661,966
; CURRENT FILING DATE: 2003-09-11
; PRIOR APPLICATION NUMBER: 60/410,595
; PRIOR FILING DATE: 2002-09-11
; PRIOR APPLICATION NUMBER: 60/422,344
; PRIOR FILING DATE: 2002-10-29
; NUMBER OF SEQ ID NOS: 61
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 19
; LENGTH: 813
; TYPE: PRT
; ORGANISM: C. elegans
US-10-661-966-19

Query Match          5.8%; Score 99.5; DB 9; Length 813;
Best Local Similarity 21.4%; Pred. No. 0.66;
Matches 55; Conservative 43; Mismatches 102; Indels 57; Gaps 15;

Qy 25 VDTRRPVP-----LTPRPVPSYAVGRGARGARLPRLRGGAHGT----QKKNQSRPQ 72
```



GenCore version 5.1.8  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: May 9, 2006, 10:55:51 ; Search time 57 Seconds  
(without alignments)  
2489.813 Million cell updates/sec

Title: US-10-634-905-2

Perfect score: 323

Sequence: 1 MKSALETRFFILLPWILIVI.....EKNLANEPKHYLDTVKIEV 323

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 2443163 seqs, 439378781 residues

Word size : 1

Total number of hits satisfying chosen parameters: 2442881

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

A\_Geneseq 21:\*

1: Geneseq1980s:\*

2: Geneseq1990s:\*

3: Geneseq2000s:\*

4: Geneseq2001s:\*

5: Geneseq2002s:\*

6: Geneseq2003as:\*

7: Geneseq2003bs:\*

8: Geneseq2004s:\*

9: Geneseq2005s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	323	100.0	323	5	ADI03930 Human enz
2	323	100.0	323	8	ADQ97589 Human can
3	197	61.0	197	5	ADI03933 HumanUDP-
4	81	25.1	145	4	AAM99900 Human exc
5	81	25.1	145	4	AAM43699 Human bla
6	81	25.1	145	8	ADF71658 Human bla
7	54	16.7	292	8	ADQ97586 Mouse can
8	51	15.8	59	3	AAB45114 Human sec
9	47	14.6	324	5	ADI03932 Rat galac
10	28	8.7	29	3	AAB45116 Human sec
11	15	4.6	334	6	ABO07264 Human p53
12	15	4.6	347	2	AAy08214 Rat glucu
13	15	4.6	347	2	AAW90057 Rat GlCAT
14	15	4.6	347	5	ABB97603 Novel hum
15	13	4.0	40	8	ABOS9511 Human gen
16	12	3.7	167	4	ABG13553 Novel hum
17	11	3.4	22	2	AAW90060 Rat GlCAT
18	11	3.4	241	4	AAW93879 Human pol
19	11	3.4	241	8	ADL31964 Human pro
20	11	3.4	335	6	ABO07265 Human p53
21	11	3.4	335	8	ADQ31194 Human GT
22	11	3.4	335	8	ADQ21418 Human sof
23	11	3.4	335	9	ADY70386 Galactosy
24	11	3.4	691	4	ABG00332 Novel hum

25	10	3.1	59	3	AAB45113	Ab45113 Human sec
26	10	3.1	201	8	ADO57267	Ado57267 Kidney de
27	10	3.1	306	8	ADI00832	Adi00832 Fruit fly
28	9	2.8	29	3	AAB45115	Ab45115 Human sec
29	8	2.5	25	2	AAW90059	Aaw90059 Rat GlCAT
30	8	2.5	107	6	ADA34275	Ada34275 Acinetoba
31	8	2.5	144	6	ABP97828	Abp97828 Amino aci
32	8	2.5	256	7	ABO76924	Abp76924 Pseudomon
33	8	2.5	272	1	AAW93560	Aap93560 Plasmodiu
34	8	2.5	316	8	ADI00836	Adi00836 Fruit fly
35	8	2.5	405	7	ABO68141	Abp68141 Pseudomon
36	8	2.5	410	7	ABO81824	Abp81824 Pseudomon
37	8	2.5	479	4	ABB62172	Abb62172 Drosophil
38	8	2.5	487	8	ADP08279	Adp08279 Neisseria
39	8	2.5	498	6	ABP79238	Abp79238 N. gonorr
40	8	2.5	530	7	ABO74593	Abp74593 Pseudomon
41	8	2.5	696	6	ABU08785	Abu08785 Transposa
42	8	2.5	912	8	ADJ76291	Adj76291 Marker ge
43	7	2.2	9	4	AAM53158	Aam53158 Human con
44	7	2.2	12	4	AAM00395	Aam00395 Human pro
45	7	2.2	19	4	AAM19546	Aam19546 Peptide #

## ALIGNMENTS

RESULT 1

ADI03930  
ID ADI03930 standard; protein; 323 AA.

XX  
AC ADI03930;

DT 22-APR-2004 (first entry)

XX  
DE Human enzyme protein.

XX  
KW Enzyme protein; drug screening; disease diagnosis; human; gene therapy;

KW chromosome 6; enzyme; glucuronyltransferase.

XX  
OS Homo sapiens.

XX  
PN WO200268657-A2.

XX  
PD 06-SEP-2002.

XX  
PF 08-FEB-2002; 2002WO-US003623.

XX  
PR 26-FEB-2001; 2001US-0270871P.

XX  
PR 26-MAR-2001; 2001US-00816095.

XX  
(PEKE ) PE CORP NY.

XX  
Gan W, Yan C, Merkulov GV, Di Francesco V, Beasley EM;

XX  
WPI; 2002-713380/77.

XX  
N-PSDB; ADI03929, ADI03931.

XX  
PT New human enzyme proteins, useful for treating or diagnosing disorders associated with abnormal expression of the protein, in drug screening assays and pharmacogenomic analysis.

XX  
PS Claim 1; SEQ ID NO 2; 127pp; English.

XX  
CC The invention relates to a novel isolated enzyme protein and encoding polynucleotides. The protein shows a high degree of similarity to a glucuronyltransferase cloned from a rabbit brain cDNA library. The peptides and nucleic acid molecules are useful as models for the development of human therapeutic targets, aid in the identification of therapeutic proteins, and serve as targets for the development of human therapeutic agents. The peptide may be used in drug screening assays, in assays to determine the biological activity of the protein, to raise antibodies or to elicit another immune response, as a reagent in assays designed to quantitatively determine levels of the protein in biological

CC fluids, or as markers for tissues in which the corresponding protein is  
CC preferentially expressed. The human enzyme protein is also useful for  
CC diagnosing a disease, predisposition to a disease, or treating a disorder  
CC characterized by an absence of, inappropriate or unwanted expression of  
CC the protein. The antibodies are useful in pharmacogenomic analysis, for  
CC inhibiting protein function, or for tissue typing. The nucleic acid  
CC molecules are useful as probes, primers, chemical intermediates, or in  
CC biological assays. The present sequence represents a human enzyme  
CC protein.  
XX  
SQ Sequence 323 AA;  
Query Match 100.0%; Score 323; DB 5; Length 323;  
Best Local Similarity 100.0%; Pred. No. 2.7e-315;  
Matches 323; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 MKSALFTRFFILLPWILIVIMLDVTRRPVPLTPRPYFSPYAVGCGARLPLRGGPA 60  
DB 1 MKSALFTRFFILLPWILIVIMLDVTRRPVPLTPRPYFSPYAVGCGARLPLRGGPA 60  
QY 61 HGQKQNSRQPQPEPQLPTIYAITPTYSRPVQKAELTRLANTFRQVAQLHWILVEDAA 120  
DB 61 HGQKQNSRQPQPEPQLPTIYAITPTYSRPVQKAELTRLANTFRQVAQLHWILVEDAA 120  
QY 121 ARSELVSRFLARAGLPSHLHVTPRRYKRPGLPRATEQORNAGLAWLRQHHQRAQPGV 180  
DB 121 ARSELVSRFLARAGLPSHLHVTPRRYKRPGLPRATEQORNAGLAWLRQHHQRAQPGV 180  
QY 181 LFFADDDNTYSLELFQEMRTTRKVSVPVGLVGGRRYERPLVGVGMYTGWADRPF 240  
DB 181 LFFADDDNTYSLELFQEMRTTRKVSVPVGLVGGRRYERPLVGVGMYTGWADRPF 240  
QY 241<sup>10</sup> AIDMAGFAVSLQVILSNPKAVFKRRGSGQGMQSDFLKQITTVVELEPKANNCTKVLVWH 300  
DB 241 AIDMAGFAVSLQVILSNPKAVFKRRGSGQGMQSDFLKQITTVVELEPKANNCTKVLVWH 300  
QY 301 TRTEKVNLANEPKYHLDTVKIEV 323  
DB 301 TRTEKVNLANEPKYHLDTVKIEV 323

RESULT 2  
ADQ97589  
ID ADQ97589 standard; protein; 323 AA.  
XX  
AC ADQ97589;  
XX  
DT 07-OCT-2004 (first entry)  
XX  
DE Human cancer associated sequence HP10-009, SEQ ID 566.  
XX  
KW Cytostatic; Gene Therapy; cancer; leukemia; lymphoma; Human.  
XX  
OS Homo sapiens.  
XX  
FN WO2004060304-A2.  
XX  
PD 22-JUL-2004.  
XX  
PF 22-DEC-2003; 2003WO-US041389.  
XX  
PR 27-DEC-2002; 2002US-00330773.  
XX  
PA (SAGR-) SAGRES DISCOVERY INC.  
XX  
PI Morris DW, Malandro MS;  
XX  
DR WPI; 2004-543781/52.  
XX  
New isolated cancer associated nucleic acids comprising at least 10  
PT contiguous nucleotides, useful for diagnosing, preventing and/or treating  
PT cancers such as leukemia and lymphoma.  
XX

PS Claim 1; SEQ ID NO 566; 199pp; English.  
XX  
CC The present invention relates to cancer associated sequences (ADQ97025-  
CC ADQ98004). The sequences are useful for the diagnosis, prevention and/or  
CC treatment of cancer, such as leukemia and lymphoma. Note: The sequence  
CC data for this patent did not form part of the printed specification, but  
CC was obtained in electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 323 AA;  
Query Match 100.0%; Score 323; DB 8; Length 323;  
Best Local Similarity 100.0%; Pred. No. 2.7e-315;  
Matches 323; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 MKSALFTRFFILLPWILIVIMLDVTRRPVPLTPRPYFSPYAVGCGARLPLRGGPA 60  
DB 1 MKSALFTRFFILLPWILIVIMLDVTRRPVPLTPRPYFSPYAVGCGARLPLRGGPA 60  
QY 61 HGQKQNSRQPQPEPQLPTIYAITPTYSRPVQKAELTRLANTFRQVAQLHWILVEDAA 120  
DB 61 HGQKQNSRQPQPEPQLPTIYAITPTYSRPVQKAELTRLANTFRQVAQLHWILVEDAA 120  
QY 121 ARSELVSRFLARAGLPSHLHVTPRRYKRPGLPRATEQORNAGLAWLRQHHQRAQPGV 180  
DB 121 ARSELVSRFLARAGLPSHLHVTPRRYKRPGLPRATEQORNAGLAWLRQHHQRAQPGV 180  
QY 181 LFFADDDNTYSLELFQEMRTTRKVSVPVGLVGGRRYERPLVGVGMYTGWADRPF 240  
DB 181 LFFADDDNTYSLELFQEMRTTRKVSVPVGLVGGRRYERPLVGVGMYTGWADRPF 240  
QY 241 AIDMAGFAVSLQVILSNPKAVFKRRGSGQGMQSDFLKQITTVVELEPKANNCTKVLVWH 300  
DB 241 AIDMAGFAVSLQVILSNPKAVFKRRGSGQGMQSDFLKQITTVVELEPKANNCTKVLVWH 300  
QY 301 TRTEKVNLANEPKYHLDTVKIEV 323  
DB 301 TRTEKVNLANEPKYHLDTVKIEV 323  
RESULT 3  
ADI03933  
ID ADI03933 standard; protein; 197 AA.  
XX  
AC ADI03933;  
XX  
DT 22-APR-2004 (first entry)  
XX  
DE HumanUDP-glucuronyltransferase-S fragment.  
XX  
KW Enzyme protein; drug screening; disease diagnosis; rat; gene therapy;  
KW chromosome 6; enzyme; glucuronyltransferase; UDP-glucuronyltransferase-S.  
XX  
OS Homo sapiens.  
XX  
FN WO200268657-A2.  
XX  
PD 06-SEP-2002.  
XX  
PF 08-FEB-2002; 2002WO-US003623.  
XX  
PR 26-FEB-2001; 2001US-0270871P.  
XX  
PR 26-MAR-2001; 2001US-00816095.  
XX  
PA (PEKE ) PE CORP NY.  
XX  
PI Gan W, Yan C, Merkulov GV, Di Francesco V, Beasley EM;  
XX  
DR WPI; 2002-713380/77.  
XX  
New human enzyme proteins, useful for treating or diagnosing disorders  
PT associated with abnormal expression of the protein, in drug screening  
PT assays and pharmacogenomic analysis.  
XX

```
XX Disclosure; SEQ ID NO 5; 127bp; English.
XX
XX The invention relates to a novel isolated enzyme protein and encoding
XX polynucleotides. The protein shows a high degree of similarity to a
XX glucuronyltransferase cloned from a rabbit brain cDNA library. The
XX peptides and nucleic acid molecules are useful as models for the
XX development of human therapeutic targets, aid in the identification of
XX therapeutic proteins, and serve as targets for the development of human
XX therapeutic agents. The peptide may be used in drug screening assays, in
XX assays to determine the biological activity of the protein, to raise
XX antibodies or to elicit another immune response, as a reagent in assays
XX designed to quantitatively determine levels of the protein in biological
XX fluids, or as markers for tissues in which the corresponding protein is
XX preferentially expressed. The human enzyme protein is also useful for
XX diagnosing a disease, predisposition to a disease, or treating a disorder
XX characterized by an absence of, inappropriate or unwanted expression of
XX the protein. The antibodies are useful in pharmacogenomic analysis, for
XX inhibiting protein function, or for tissue typing. The nucleic acid
XX molecules are useful as probes, primers, chemical intermediates, or in
XX biological assays. The present sequence represents a human UDP-
XX glucuronyltransferase-S fragment used in alignment studies with the novel
XX human enzyme protein.
XX
XX Sequence 197 AA;
XX
XX Query Match 61.0%; Score 197; DB 5; Length 197;
XX Best Local Similarity 100.0%; Pred. No. 5.6e-189; Indels 0; Gaps 0;
XX Matches 197; Conservative 0; Mismatches 0;
XX
Qy 1 MKSALFRFFILLPWILVIMLDVTRRPVPLTPRPYSPYAVGCGARLPLRRGGPA 60
Db 1 MKSALFRFFILLPWILVIMLDVTRRPVPLTPRPYSPYAVGCGARLPLRRGGPA 60
Qy 61 HGTKRQNSRQPQPEQLPTIYAITYSRVPQKAELTRANTFRVAQLHWILVEDAA 120
Db 61 HGTKRQNSRQPQPEQLPTIYAITYSRVPQKAELTRANTFRVAQLHWILVEDAA 120
Qy 121 ARSELVSRFLARAGLPSTLHVPTPRYKRPGLPRATEQRNAGLAWLRQHRHQRAQPGV 180
Db 121 ARSELVSRFLARAGLPSTLHVPTPRYKRPGLPRATEQRNAGLAWLRQHRHQRAQPGV 180
Qy 181 LFFADDDNTYSLELFOE 197
Db 181 LFFADDDNTYSLELFOE 197
RESULT 4
AAM99900
ID AAM99900 standard; protein; 145 AA.
AC AAM99900;
XX
XX 07-JAN-2002 (first entry)
XX
XX Human excretory related polypeptide SEQ ID NO 637.
XX
XX Human; nootropic; neuroprotective; cytostatic; dermatological; virucide;
XX immunosuppressive; antiinflammatory; anti-HIV; antibacterial; vulnerary;
XX antiparkinsonian; antisickling; antianaemic; antiarthritic; cancer;
XX antirheumatic; hepatotropic; cerebroprotective; antiinflammatory;
XX antiallergic; antidiabetic; antulcer; anticonvulsant; antifungal;
XX antiparasitic; cardiant; immune disorder; cardiovascular disorder;
XX neurological disease; infection; nephrotropic; gene therapy; vaccine;
XX excretory system.
XX
XX Homo sapiens.
XX
XX W0200155313-A2.
XX
XX 02-AUG-2001.
XX
XX 17-JAN-2001; 2001WO-US001323.
XX
```

```
XX
PR 31-JAN-2000; 2000US-0179065P. XX
PR 04-FEB-2000; 2000US-0180628P. PR
PR 24-FEB-2000; 2000US-0184664P. PR
PR 02-MAR-2000; 2000US-0186350P. PR
PR 16-MAR-2000; 2000US-0189874P. PR
PR 17-MAR-2000; 2000US-0190076P. PR
PR 18-APR-2000; 2000US-0198123P. PR
PR 19-MAY-2000; 2000US-0205515P. PR
PR 07-JUN-2000; 2000US-0209467P. PR
PR 28-JUN-2000; 2000US-0214886P. PR
PR 30-JUN-2000; 2000US-0215135P. PR
PR 07-JUL-2000; 2000US-0216647P. PR
PR 11-JUL-2000; 2000US-0216880P. PR
PR 11-JUL-2000; 2000US-0217487P. PR
PR 14-JUL-2000; 2000US-0218290P. PR
PR 26-JUL-2000; 2000US-0220963P. PR
PR 26-JUL-2000; 2000US-0220964P. PR
PR 14-AUG-2000; 2000US-0224518P. PR
PR 14-AUG-2000; 2000US-0224519P. PR
PR 14-AUG-2000; 2000US-0225213P. PR
PR 14-AUG-2000; 2000US-0225214P. PR
PR 14-AUG-2000; 2000US-0225266P. PR
PR 14-AUG-2000; 2000US-0225267P. PR
PR 14-AUG-2000; 2000US-0225268P. PR
PR 14-AUG-2000; 2000US-0225270P. PR
PR 14-AUG-2000; 2000US-0225447P. PR
PR 14-AUG-2000; 2000US-0225757P. PR
PR 14-AUG-2000; 2000US-0225758P. PR
PR 14-AUG-2000; 2000US-0225759P. PR
PR 18-AUG-2000; 2000US-0226279P. PR
PR 22-AUG-2000; 2000US-0226681P. PR
PR 22-AUG-2000; 2000US-0226868P. PR
PR 22-AUG-2000; 2000US-0227182P. PR
PR 23-AUG-2000; 2000US-0227009P. PR
PR 30-AUG-2000; 2000US-0228924P. PR
PR 01-SEP-2000; 2000US-0229287P. PR
PR 01-SEP-2000; 2000US-0229343P. PR
PR 01-SEP-2000; 2000US-0229344P. PR
PR 01-SEP-2000; 2000US-0229345P. PR
PR 05-SEP-2000; 2000US-0229509P. PR
PR 05-SEP-2000; 2000US-0229513P. PR
PR 06-SEP-2000; 2000US-0230437P. PR
PR 06-SEP-2000; 2000US-0230438P. PR
PR 08-SEP-2000; 2000US-0231242P. PR
PR 08-SEP-2000; 2000US-0231243P. PR
PR 08-SEP-2000; 2000US-0231244P. PR
PR 08-SEP-2000; 2000US-0231413P. PR
PR 08-SEP-2000; 2000US-0231414P. PR
PR 08-SEP-2000; 2000US-0232080P. PR
PR 08-SEP-2000; 2000US-0232081P. PR
PR 12-SEP-2000; 2000US-0231968P. PR
PR 14-SEP-2000; 2000US-0232397P. PR
PR 14-SEP-2000; 2000US-0232398P. PR
PR 14-SEP-2000; 2000US-0232399P. PR
PR 14-SEP-2000; 2000US-0232400P. PR
PR 14-SEP-2000; 2000US-0232401P. PR
PR 14-SEP-2000; 2000US-0233063P. PR
PR 14-SEP-2000; 2000US-0233064P. PR
PR 14-SEP-2000; 2000US-0233065P. PR
PR 21-SEP-2000; 2000US-0234223P. PR
PR 21-SEP-2000; 2000US-0234274P. PR
PR 25-SEP-2000; 2000US-0234997P. PR
PR 25-SEP-2000; 2000US-0234998P. PR
PR 26-SEP-2000; 2000US-0235484P. PR
PR 27-SEP-2000; 2000US-0235834P. PR
PR 27-SEP-2000; 2000US-0235836P. PR
PR 29-SEP-2000; 2000US-0236327P. PR
PR 29-SEP-2000; 2000US-0236367P. PR
PR 29-SEP-2000; 2000US-0236368P. PR
PR 29-SEP-2000; 2000US-0236369P. PR
PR 29-SEP-2000; 2000US-0236370P. PR
```





CC urogenital system; immune disorders such as Addison's disease, allergies,  
CC autoimmunity, haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus,  
CC Crohn's disease, multiple sclerosis, rheumatoid arthritis and ulcerative  
CC colitis; cardiovascular disorders such as myocardial ischaemias; wound  
CC healing; neurological diseases such as cerebral anoxia and epilepsy; and  
CC infectious diseases such as viral, bacterial, fungal and parasitic  
CC infections. Numerous examples of each type of disorder are given in the  
CC specification. The polypeptides can also be used as a food additive or  
CC preservative to increase or decrease storage capabilities. The  
CC polynucleotides are useful for chromosome identification. They are also  
CC useful as probes for diagnosing or treating a disorder related to the  
CC female reproductive system, particularly breast and/or ovary cancer. The  
CC present sequence is a novel bladder antigen provided in the invention.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences  
XX  
SQ Sequence 145 AA;

Query Match 25.1%; Score 81; DB 4; Length 145;  
Best Local Similarity 100.0%; Pred. No. 1.3e-72;  
Matches 81; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 210 GLVGGRRYERPLVENGKVGWYTGWRADRFALDMAGFAVSLQVILSNPKAVFKRGSGP 269  
Db 23 GLVGGRRYERPLVENGKVGWYTGWRADRFALDMAGFAVSLQVILSNPKAVFKRGSGP 82  
QY 270 GMOESDFLKQITVVELEPKA 290  
Db 83 GMOESDFLKQITVVELEPKA 103

RESULT 6  
ADF71658  
ID ADF71658 standard; protein; 145 AA.  
XX  
AC ADF71658;  
XX  
DT 11-MAR-2004 (first entry)  
XX Human bladder associated antigen #38.  
XX  
KW autoimmune disease; rheumatoid arthritis; hyperproliferative disorder;  
KW breast neoplasms; liver neoplasms; cardiovascular disorder;  
KW cardiac arrest; cerebrovascular disorder; cerebral ischaemia;  
KW angiogenesis; nervous system disorder; Alzheimer's disease; infection;  
KW ocular disorder; corneal infection; wound healing;  
KW epithelial cell proliferation; skin aging; sunburn;  
KW organ transplantation; cell culture; tissue regeneration; chemotaxis;  
KW food additive; preservative; Human; bladder associated antigen.  
XX

OS Homo sapiens.  
XX  
XX US2003199008-A1.  
XX 23-OCT-2003.  
XX  
XX 22-FEB-2002; 2002US-00080254.  
XX  
XX 31-JAN-2000; 2000US-0179065P.  
PR 04-FEB-2000; 2000US-0180628P.  
PR 24-FEB-2000; 2000US-0184664P.  
PR 02-MAR-2000; 2000US-0186350P.  
PR 16-MAR-2000; 2000US-0189874P.  
PR 17-MAR-2000; 2000US-0190076P.  
PR 18-APR-2000; 2000US-0198123P.  
PR 19-MAY-2000; 2000US-0205515P.  
PR 07-JUN-2000; 2000US-0209467P.  
PR 28-JUN-2000; 2000US-0214886P.  
PR 30-JUN-2000; 2000US-0215135P.  
PR 07-JUL-2000; 2000US-0216647P.  
PR 07-JUL-2000; 2000US-0216880P.  
PR 11-JUL-2000; 2000US-0217487P.

PR 11-JUL-2000; 2000US-0217496P.  
PR 14-JUL-2000; 2000US-0218290P.  
PR 26-JUL-2000; 2000US-0220963P.  
PR 26-JUL-2000; 2000US-0220964P.  
PR 14-AUG-2000; 2000US-0224518P.  
PR 14-AUG-2000; 2000US-0224519P.  
PR 14-AUG-2000; 2000US-0225213P.  
PR 14-AUG-2000; 2000US-0225214P.  
PR 14-AUG-2000; 2000US-0225266P.  
PR 14-AUG-2000; 2000US-0225267P.  
PR 14-AUG-2000; 2000US-0225268P.  
PR 14-AUG-2000; 2000US-0225270P.  
PR 14-AUG-2000; 2000US-0225447P.  
PR 14-AUG-2000; 2000US-0225757P.  
PR 14-AUG-2000; 2000US-0225758P.  
PR 14-AUG-2000; 2000US-0225759P.  
PR 18-AUG-2000; 2000US-0226279P.  
PR 22-AUG-2000; 2000US-0226681P.  
PR 22-AUG-2000; 2000US-0226868P.  
PR 22-AUG-2000; 2000US-0227182P.  
PR 23-AUG-2000; 2000US-0227009P.  
PR 30-AUG-2000; 2000US-0228924P.  
PR 01-SEP-2000; 2000US-0229287P.  
PR 01-SEP-2000; 2000US-0229343P.  
PR 01-SEP-2000; 2000US-0229344P.  
PR 01-SEP-2000; 2000US-0229345P.  
PR 05-SEP-2000; 2000US-0229509P.  
PR 05-SEP-2000; 2000US-0229513P.  
PR 06-SEP-2000; 2000US-0230437P.  
PR 06-SEP-2000; 2000US-0230438P.  
PR 08-SEP-2000; 2000US-0231242P.  
PR 08-SEP-2000; 2000US-0231243P.  
PR 08-SEP-2000; 2000US-0231244P.  
PR 08-SEP-2000; 2000US-0231413P.  
PR 08-SEP-2000; 2000US-0231414P.  
PR 08-SEP-2000; 2000US-0232080P.  
PR 08-SEP-2000; 2000US-0232081P.  
PR 12-SEP-2000; 2000US-0231968P.  
PR 14-SEP-2000; 2000US-0232397P.  
PR 14-SEP-2000; 2000US-0232398P.  
PR 14-SEP-2000; 2000US-0232399P.  
PR 14-SEP-2000; 2000US-0232400P.  
PR 14-SEP-2000; 2000US-0232401P.  
PR 14-SEP-2000; 2000US-0233063P.  
PR 14-SEP-2000; 2000US-0233064P.  
PR 14-SEP-2000; 2000US-0233065P.  
PR 21-SEP-2000; 2000US-0234223P.  
PR 21-SEP-2000; 2000US-0234274P.  
PR 25-SEP-2000; 2000US-0234997P.  
PR 25-SEP-2000; 2000US-0234998P.  
PR 26-SEP-2000; 2000US-0235484P.  
PR 27-SEP-2000; 2000US-0235834P.  
PR 27-SEP-2000; 2000US-0235836P.  
PR 29-SEP-2000; 2000US-0236327P.  
PR 29-SEP-2000; 2000US-0236367P.  
PR 29-SEP-2000; 2000US-0236368P.  
PR 29-SEP-2000; 2000US-0236369P.  
PR 29-SEP-2000; 2000US-0236370P.  
PR 02-OCT-2000; 2000US-0236802P.  
PR 02-OCT-2000; 2000US-0237037P.  
PR 02-OCT-2000; 2000US-0237038P.  
PR 02-OCT-2000; 2000US-0237039P.  
PR 02-OCT-2000; 2000US-0237040P.  
PR 13-OCT-2000; 2000US-0239935P.  
PR 13-OCT-2000; 2000US-0239937P.  
PR 20-OCT-2000; 2000US-0240960P.  
PR 20-OCT-2000; 2000US-0241221P.  
PR 20-OCT-2000; 2000US-0241785P.  
PR 20-OCT-2000; 2000US-0241786P.  
PR 20-OCT-2000; 2000US-0241787P.  
PR 20-OCT-2000; 2000US-0241808P.  
PR 20-OCT-2000; 2000US-0241809P.  
PR 20-OCT-2000; 2000US-0241826P.





Best Local Similarity	100.0%;	Pred. No. 2.9e-45;	Matches 54;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	237	DRPFAIDMAGFAVLSQVILSNPKAVFKRGSGPGMQESDFLKQITTTVELEPKA	290				
Db	206	DRPFAIDMAGFAVLSQVILSNPKAVFKRGSGPGMQESDFLKQITTTVELEPKA	259				
RESULT 8							
ID	AAB45114						
AC	AAB45114						
XX							
DT	12-FEB-2001	(first entry)					
DE	Human secreted protein encoded by gene 46 homologue.						
XX							
DE	Secreted protein; human; immunosuppressive; antiarthritic; antirheumatic;						
KW	antiproliferative; cytostatic; cardiant; vasotropic; cerebroprotective;						
KW	nootropic; neuroprotective; antibacterial; virucide; fungicide;						
KW	ophthalmological; vulnery; gene therapy; treatment; autoimmune disease;						
KW	hyperproliferative disorder; cardiovascular disorder; ocular disorder;						
KW	cerebrovascular disorder; nervous system disorder; infection; skin aging;						
KW	wound healing; epithelial cell proliferation; transplantation.						
XX							
OS	Homo sapiens.						
XX							
PN	WO200058357-A1.						
XX							
PD	05-OCT-2000.						
XX							
PF	23-MAR-2000;	2000WO-US0007723.					
XX							
PR	26-MAR-1999;	99US-0126506P.					
PR	07-JAN-2000;	2000US-0174852P.					
XX							
PA	(HUMA-) HUMAN GENOME SCI INC.						
XX							
PI	Rosen CA, Ruben SM, Komatsoulis G;						
XX							
DR	WPI; 2000-611704/58.						
XX							
PT	Nucleic acid molecules encoding human secreted proteins, used in						
PT	preventing, treating or ameliorating a disorder, e.g. Alzheimer's and						
PT	Parkinson's diseases and cancers.						
XX							
PS	Disclosure; Page 413; 418pp; English.						
XX							
CC	This invention describes novel isolated nucleic acid molecules (I)						
CC	encoding a human secreted proteins (II) which have immunosuppressive,						
CC	antiarthritic, antirheumatic, antiproliferative, cytostatic, cardiant,						
CC	vasotropic, cerebroprotective, nootropic, neuroprotective, antibacterial,						
CC	virucide, fungicide, ophthalmological and vulnery activity and can be						
CC	used for gene therapy. (I) and (II) are used to prevent, treat or						
CC	ameliorate a medical condition in e.g. humans, mice, rabbits, goats,						
CC	horses, cats, dogs, chickens or sheep. (I) and (II) are also used in						
CC	diagnosing a pathological condition or susceptibility to a pathological						
CC	condition. The antibodies to (II) can also be used in alleviating						
CC	symptoms associated with the disorders and in diagnostic immunoassays						
CC	e.g. radioimmunoassays or enzyme linked immunosorbent assays (ELISA).						
CC	Disorders which are diagnosed or treated include autoimmune diseases e.g.						
CC	rheumatoid arthritis, hyperproliferative disorders e.g. neoplasms of the						
CC	breat or liver, cardiovascular disorders e.g. cardiac arrest,						
CC	cerebrovascular disorders e.g. cerebral ischemia, angiogenesis, nervous						
CC	system disorders e.g. Alzheimer's disease, infections caused by bacteria,						
CC	viruses and fungi and ocular disorders e.g. corneal infection. The						
CC	polypeptides can also be used to aid wound healing and epithelial cell						
CC	proliferation, to prevent skin aging due to sunburn, to maintain organs						
CC	before transplantation, for supporting cell culture of primary tissues,						
CC	to regenerate tissues and in chemotaxis. The polypeptides can also be						
CC	used as a food additive or preservative to increase or decrease storage						
CC	capabilities						
XX							
Sequence 59 AA;							
Query Match	15.8%;	Score 51;	DB 3;	Length 59;			
Best Local Similarity	100.0%;	Pred. No. 7.7e-43;					
Matches 51;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;			
QY	246	GFAVLSQVILSNPKAVFKRGSGPGMQESDFLKQITTTVELEPKANNCTKV	296				
Db	7	GFAVLSQVILSNPKAVFKRGSGPGMQESDFLKQITTTVELEPKANNCTKV	57				
RESULT 9							
AD103932							
ID	AD103932	standard; protein; 324 AA.					
XX							
AC	AD103932;						
XX							
DT	22-APR-2004	(first entry)					
DE	Rat galactosidase beta-1,3-glucuronyltransferase.						
XX							
KW	Enzyme protein; drug screening; disease diagnosis; rat; gene therapy;						
KW	chromosome 6; enzyme; glucuronyltransferase;						
KW	galactosidase beta-1,3-glucuronyltransferase.						
XX							
OS	Rattus norvegicus.						
XX							
PN	WO200268657-A2.						
XX							
PD	06-SEP-2002.						
XX							
PF	08-FEB-2002;	2002WO-US003623.					
XX							
PR	26-FEB-2001;	2001US-0270871P.					
PR	26-MAR-2001;	2001US-00816095.					
XX							
PA	(PEKE ) PE CORP NY.						
XX							
PI	Can W, Yan C, Merkulov GV, Di Francesco V, Beasley EM;						
XX							
DR	WPI; 2002-713380/77.						
DR	REFSEQ; NP_072131.1.						
XX							
PT	New human enzyme proteins, useful for treating or diagnosing disorders						
PT	associated with abnormal expression of the protein, in drug screening						
PT	assays and pharmacogenomic analysis.						
XX							
PS	Disclosure; SEQ ID NO 4; 127pp; English.						
XX							
CC	The invention relates to a novel isolated enzyme protein and encoding						
CC	polynucleotides. The protein shows a high degree of similarity to a						
CC	glucuronyltransferase cloned from a rabbit brain cDNA library. The						
CC	peptides and nucleic acid molecules are useful as models for the						
CC	development of human therapeutic targets, aid in the identification of						
CC	therapeutic proteins, and serve as targets for the development of human						
CC	therapeutic agents. The peptide may be used in drug screening assays, in						
CC	assays to determine the biological activity of the protein, to raise						
CC	antibodies or to elicit another immune response, as a reagent in assays						
CC	designed to quantitatively determine levels of the protein in biological						
CC	fluids, or as markers for tissues in which the corresponding protein is						
CC	preferentially expressed. The human enzyme protein is also useful for						
CC	diagnosing a disease, predisposition to a disease, or treating a disorder						
CC	characterized by an absence of, inappropriate or unwanted expression of						
CC	the protein. The antibodies are useful in pharmacogenomic analysis, for						
CC	inhibiting protein function, or for tissue typing. The nucleic acid						
CC	molecules are useful as probes, primers, chemical intermediates, or in						
CC	biological assays. The present sequence represents a rat galactosidase						
CC	beta-1,3-glucuronyltransferase used in alignment studies with the novel						
XX	human enzyme protein.						
XX							
Sequence 324 AA;							



CC Modulators identified by (M1) are useful in a variety of diagnostic and  
 CC therapeutic applications, where disease or disorder prognosis is related  
 CC to defects in the p53 pathway, such as, angiogenesis, apoptotic or cell  
 CC proliferation disorders (e.g. cancer). Another two new methods (M2 and  
 CC M3) are useful for modulating the p53 pathway of a cell, thus restoring  
 CC the p53 function of the cell, so that the cell undergoes normal  
 CC proliferation or progression through the cell cycle. (M2) and (M3) are  
 CC also useful for treating defects in the p53 pathway such as angiogenic,  
 CC apoptotic or cell proliferation disorders. The present sequence  
 CC represents a human p53 pathway modifying protein  
 XX  
 SQ Sequence 334 AA;

Query Match 4.6%; Score 15; DB 6; Length 334;  
 Best Local Similarity 100.0%; Pred. No. 4.7e-06;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 86 TPTYSRPVQKAELTR 100  
 Db 90 TPTYSRPVQKAELTR 104  
 |||||

RESULT 12  
 AAY08214  
 ID AAY08214 standard; protein; 347 AA.

XX AC AAY08214;  
 XX DT 09-JUL-1999 (first entry)  
 XX DE Rat glucuronyltransferase protein.

XX HNK-1; sulphotransferase; rat; enzyme; carbohydrate epitope; nCAM;  
 KW neural cell adhesion molecule; neural growth; remyelination; disease;  
 KW neural protection; treatment; neuron; glucuronyltransferase.

XX OS Rattus sp.  
 XX PN WO9911796-A1.

XX PD 11-MAR-1999.  
 XX PF 04-SEP-1998; 98WO-US018572.

XX PR 05-SEP-1997; 97US-00924812.  
 XX PA (ACOR-) ACORDA THERAPEUTICS.

XX PI Mantei N, Bakker H, Schachner M;

XX DR WPI; 1999-312405/26.  
 XX DR N-PSDB; AAX37750.

XX PT New enzyme useful for promoting neural growth, remyelination and neural  
 PT protection.

XX PS Disclosure; Page 80-82; 85pp; English.

XX CC This invention describes a novel isolated rat sulphotransferase enzyme  
 CC capable of generating the HNK-1 carbohydrate epitope on a neural cell  
 CC adhesion molecule (nCAM). HNK-1 sulphotransferase is used to promote  
 CC neural growth, remyelination, and/or neural protection in mammals. The  
 CC cDNA encoding the HNK-1 sulphotransferase enzyme is useful in the  
 CC treatment of diseased or damaged neurons. The HNK-1 sulphotransferase  
 CC cDNA and protein are also useful in screening for potential drugs  
 CC effective to modulate the sulphotransferase activity of target mammalian  
 CC cells. The product of the invention has neuroprotective activity. This  
 CC sequence represents a rat glucuronyltransferase protein used to describe  
 CC the method of the invention

XX SQ Sequence 347 AA;

Query Match 4.6%; Score 15; DB 2; Length 347;

Best Local Similarity 100.0%; Pred. No. 4.8e-06;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 86 TPTYSRPVQKAELTR 100  
 Db 103 TPTYSRPVQKAELTR 117  
 |||||

RESULT 13  
 AAW90057  
 ID AAW90057 standard; protein; 347 AA.

XX AC AAW90057;

XX DT 26-FEB-1999 (first entry)

XX DE Rat GlcAT-P protein.

XX KW GlcAT-P; glucuronate transferase; rat; glucuronic acid; donor; receptor.

XX OS Rattus sp.

XX PN JP10313867-A.

XX PD 02-DEC-1998.

XX PF 16-MAY-1997; 97JP-00127065.

XX PR 16-MAY-1997; 97JP-00127065.

XX PA (SEGK) SEIKAGAKU KOGYO CO LTD.

XX DR WPI; 1999-074149/07.

XX DR N-PSDB; AAV73869.

XX PT DNA coding glucuronate transferase - comprises transferring glucuronic  
 PT acid from glucuronic acid donor to glucuronic acid receptor.

XX PS Claim 3; Page 13-14; 18pp; Japanese.

XX CC This sequence represents a novel rat glucuronate transferase (GlcAT-P)  
 CC which transfers glucuronic acid from a glucuronic acid donor to a  
 CC glucuronic acid receptor with an optimum reaction pH of 6.0 to 6.5. The  
 CC protein transfers specifically glucuronic acid to the N-  
 CC acetylglucosamine residue of asialo-orosomucoid and neuron adhesive  
 CC molecule, has a molecular weight of approximately 45,000 and maintains  
 CC its maintains its activity in the presence of 5 mM neolactotetraose-  
 CC phenylC14H29

XX SQ Sequence 347 AA;

Query Match 4.6%; Score 15; DB 2; Length 347;

Best Local Similarity 100.0%; Pred. No. 4.8e-06;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 86 TPTYSRPVQKAELTR 100  
 Db 103 TPTYSRPVQKAELTR 117  
 |||||

RESULT 14  
 ABB97603  
 ID ABB97603 standard; protein; 347 AA.

XX AC ABB97603;

XX DT 27-JUN-2002 (first entry)

XX DE Novel human protein SEQ ID NO: 871.

XX KW Human; antianaemic; vulnery; antiinflammatory; immunomodulator;  
 KW antinfertility; cerebroprotective; cytostatic; rheumatic; gene therapy;  
 KW neuroprotective; antiparkinsonian; protein therapy; EST;

```
KW expressed sequence tag.
XX Homo sapiens.
XX WO200222660-A2.
XX 21-MAR-2002.
XX 10-SEP-2001; 2001WO-US026015.
XX 11-SEP-2000; 2000US-00659671.
XX (HYSE-) HYSEQ INC.
XX Tang YT, Liu C, Zhou P, Asundi V, Zhang J, Zhao QA, Ren F;
XX Xue AJ, Yang Y, Wehrman T, Drmanac RT;
XX WPI; 2002-292408/33.
XX N-PSDB; ABN32789.
XX An isolated polynucleotide for treating diseases associated with its
XX encoded polypeptide such as cancer and multiple sclerosis.
XX Example 2; SEQ ID NO 871; 509pp; English.
XX The present invention provides the protein and coding sequences of 444
XX novel human proteins. These were isolated from expressed sequences tags
XX (ESTs). They can be used to stimulate cell growth, to regulate
XX haematopoiesis e.g. to treat aplastic anaemia, to help tissue regrowth
XX e.g. in burn treatment, to regulate the immune system e.g. to treat
XX multiple sclerosis, to regulate activin or inhibit e.g. to treat
XX infertility, to regulate haemostasis or thrombolysis e.g. to treat stroke
XX and cancer, to screen for drugs, to treat inflammatory conditions e.g.
XX rheumatoid arthritis, and to treat nervous system disorders e.g.
XX Parkinson's disease. The present sequence is a protein of the invention
XX SQ Sequence 347 AA;
Query Match 4.6%; Score 15; DB 5; Length 347;
Best Local Similarity 100.0%; Pred. No. 4.8e-06;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 86 TPTYSRPVQKALTR 100
Db 103 TPTYSRPVQKALTR 117
RESULT 15
ABO59511
ID ABO59511 standard; protein; 40 AA.
XX AC ABO59511;
XX 29-JUL-2004 (first entry)
XX Human genome derived single exon protein #5745.
XX DE Human; gene expression; single exon probe; microarray;
XX KW alternative splicing event; genomic alteration.
XX OS Homo sapiens.
XX PN US2003194704-A1.
XX 16-OCT-2003.
XX 03-APR-2002; 2002US-00029386.
XX 03-APR-2002; 2002US-00029386.
XX (PENN/) PENN S G.
XX PA (RANK/) RANK D R.
XX PA (HANZ/) HANZEL D K.
```

```
XX Penn SG, Rank DR, Hanzel DK;
XX WPI; 2004-119264/12.
XX New human genome-derived single exon nucleic acid probes useful for human
XX gene expression analysis, for identifying or characterizing alternative
XX splicing events, for assessing genomic alterations or as tools for
XX surveying tissues.
XX Claim 45; SEQ ID NO 33145; 80pp; English.
XX The invention relates to a nucleic acid probe for measuring human gene
XX expression, comprising any of the 27,400 fully defined nucleotide
XX sequences in the specification, or their complements or fragments, and
XX encoding at least 8 amino acids of any of the 6888 amino acid sequences
XX fully defined in the specification. The probe is a single exon probe that
XX hybridises under high stringency conditions to a nucleic acid molecule
XX expressed in human cells or tissues. Also included are a spatially-
XX addressable set of single exon nucleic acid probes for measuring human
XX gene expression (comprising a plurality of single exon nucleic acid
XX probes cited above, where each of the plurality of probes is separately
XX and addressably isolatable or amplifiable from the plurality), a single
XX exon microarray for measuring human gene expression, a method of
XX measuring human gene expression, a vector comprising the single exon
XX probe cited above, an ORF-encoded peptide comprising at least 8
XX contiguous amino acids of any of the above-mentioned amino acid
XX sequences (optionally with conservative amino acid substitutions), an
XX isolated antibody that binds specifically to a peptide cited above,
XX methods of selling and/or licensing single exon probes or microarrays to
XX a customer desiring to measure gene expression, a method of providing
XX human gene expression data by subsequence, and a computer-readable
XX storage medium which contains a database having a plurality of records
XX (each record including data on the expression of a single exon probe
XX cited above. The probe, methods and apparatus are useful in gene
XX expression analysis. The probes may be used as tools for surveying
XX tissues to detect the presence of expressed messages that contain their
XX specific exon, or in constructing genome-derived single exon microarrays.
XX In addition, the probes are used in identifying and characterising
XX alternative splicing events, in detecting and characterising gross
XX alterations in the genomic locus that includes their exon, in assessing
XX smaller genomic alterations, in priming the synthesis of nucleic acids,
XX or in expressing the ORF-encoded peptide. The present sequence is a human
XX single exon probe protein of the invention. Note: The sequence data for
XX this patent did not form part of the printed specification, but was
XX obtained in electronic format directly from USPTO at
XX seqdata.uspto.gov/sequence.html?DocID=20030194704
XX SQ Sequence 40 AA;
Query Match 4.0%; Score 13; DB 8; Length 40;
Best Local Similarity 100.0%; Pred. No. 7.8e-05;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 183 FADDDNTYSLELF 195
Db 26 FADDDNTYSLELF 38
Search completed: May 9, 2006, 10:57:06
Job time : 58 secs
```

**This Page Blank (uspro)**